

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 21, 2004, 13:00:37 ; Search time 34.5 Seconds
(without alignments)
11158.217 Million cell updates/sec

Title: US-10-029-345A-108_COPY_532_2532

Perfect score: 667
Sequence: 1 gttgctcatgcccacatgagat.....tggaaatcatctgagctcc 2001

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Word size: 1

Total number of hits satisfying chosen parameters: 115352

Minimum DB seq length: 473

Maximum DB seq length: 200000000

Post-processing: Listing first 75 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xip
-O/cgrr 1/USPTO.spool.p/US10029345/runat_21062004_124616_5706/app_query.fasta_1.2183
-DB=PIR 78 -OPMT=faaetan -SUFFIX=oligo.rpr -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdt -LIST=75
-DOCALLGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=50 -MODE=LOCAL -OUTFMT=ptio
-NORM=ext -HEAPSIZE=500 -MINLEN=473 -MAXLEN=200000000
-USER=US10029345_@CGN 1.1.52 @runat_21062004_124616_5706 -NCPU=6 -ICPU=3
-NO MMP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAEXT=60 -FGAPOP=6
-FGAEXT=7 -YGAPOP=60 -YGAEXT=60 -DELOP=6 -DELEXT=7

Database :

PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	1.3	619	2 T15969	hypothetical prote
2	9	1.4	737	2 A34402	peptidyl-di-peptida
3	9	1.4	890	2 A97750	DNA mismatch repair
4	8	1.2	480	1 IBEG	hydroxymethylbilan
5	8	1.2	515	2 G86332	FeF9.1 protein - A
6	8	1.2	525	2 E96786	protein F10A5.13 [
7	8	1.2	526	1 P5XR10	outer capsid prote
8	8	1.2	545	2 P95325	probable arylsulfa
9	8	1.2	555	2 VGN211	cell fusion glycop
10	8	1.2	605	1 A35459	glucose oxidase (E
11	8	1.2	640	1 QOBEU2	U135 protein - hum
12	8	1.2	672	2 F84524	En/Spm-like transp
13	8	1.2	677	2 T22333	histidine ammonia-
14	8	1.2	692	2 T52120	acyl-CoA oxidase (

15	8	1.2	727	2 AB2353	hypothetical prote
16	8	1.2	752	2 H90135	cell division cycl
17	8	1.2	789	2 AE2688	Na+/H+ antiporter
18	8	1.2	789	2 H97469	probable NADH dehy
19	8	1.2	835	2 A27211	vira protein - Agr
20	8	1.2	1153	2 T21386	hypothetical prote
21	8	1.2	1286	2 T18734	hypothetical prote
22	8	1.2	1308	2 T15280	hypothetical prote
23	8	1.2	1716	2 T14103	probable DNA-direc
24	8	1.2	1717	2 T13961	DNA-directed RNA p
25	8	1.2	2512	1 MNWV82	nonstructural poly
26	8	1.2	2514	1 T29285	nonstructural poly
27	8	1.2	2761	2 RBH7AP	adenomalous polyp
28	8	1.2	2843	1 RBH7AP	adenomalous polyp
29	8	1.2	2845	2 T49505	adenomalous polyp
30	8	1.2	473	2 B97073	beta-glucosidase [
31	7	1.1	474	2 D91075	6-phospho-beta-glu
32	7	1.1	474	2 C85920	6-phospho-beta-glu
33	7	1.1	474	2 H65051	6-phospho-beta-glu
34	7	1.1	475	2 S49886	probable membrane
35	7	1.0	475	2 B70143	fibronectin/fibrin
36	7	1.0	476	2 H89773	hypothetical prote
37	7	1.1	477	2 E97073	phospho-beta-gluco
38	7	1.1	477	2 G83048	probable two-compo
39	7	1.1	478	2 C86807	beta-glucosidase (
40	7	1.0	478	2 S17836	fatty-acyl-CoA red
41	7	1.1	479	2 F70573	hypothetical prote
42	7	1.0	479	2 T47415	transporter-like p
43	7	1.1	480	2 H86411	protein F1K23.12 [
44	7	1.1	481	2 H69593	6-phospho-beta-glu
45	7	1.1	482	2 AB1421	beta-glucosidase h
46	7	1.1	483	1 A31521	orphan receptor TR
47	7	1.0	483	2 F87700	rfa protein [limp
48	7	1.1	484	2 B33501	myosin heavy chain
49	7	1.1	486	2 AF1107	phospho-beta-gluco
50	7	1.1	486	2 A11468	phospho-beta-gluco
51	7	1.1	486	2 D89911	amino acid carrier
52	7	1.0	486	2 S67589	probable membrane
53	7	1.1	487	1 T32635	phosphoprotein pho
54	7	1.0	488	2 AE0001	conserved hypochet
55	7	1.0	489	1 VGXEMV	surface glycoprote
56	7	1.1	490	2 T02545	probable MYB fami
57	7	1.1	490	2 A24262	phase-1 i flagell
58	7	1.1	491	2 S49790	probable membrane
59	7	1.0	493	2 T43465	hypothetical prote
60	7	1.1	494	2 S16121	flagellin - Salmon
61	7	1.1	496	2 S55665	hypothetical prote
62	7	1.0	496	2 T19776	Na+/H+ antiporter
63	7	1.0	498	2 F89861	probable kinase pr
64	7	1.1	499	2 AF0073	hypothetical prote
65	7	1.1	501	2 F87550	hypothetical prote
66	7	1.0	502	2 H75290	hypothetical prote
67	7	1.1	504	2 F72744	protein F5011.6 [i
68	7	1.1	505	2 B86258	probable aspartic
69	7	1.0	506	2 T07915	hypothetical prote
70	7	1.0	506	2 F86253	probable WD-40 rep
71	7	1.0	507	2 G84577	probable long chai
72	7	1.1	508	2 H95271	cytochrome (EC 3.4.2
73	7	1.0	509	2 S49349	conserved hypochet
74	7	1.0	509	2 D86911	glyceraldehyde 3-p
75	7	1.1	509	2 G82409	

ALIGNMENTS

RESULT 1
T15969
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #ext_change 20-Sep-1999
C:Accession: T15969
R:Chisase, S.
submitted to the EMBL Data Library, July 1995

A:Description: The sequence of *C. elegans* cosmid F08B1.
A:Reference number: Z18439
A:Accession: J15969
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-619 <CH>
A:Cross-references: EMBL:U23178; NID:G726421; PID:G726422; PIDN:AA046719.1; CESP:F08B1.1
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:F08B1.1
A:Introns: 22/2; 57/1; 99/1; 160/3; 252/2; 549/3

Alignment Scores:
Pred. No.: 4.76 Length: 619
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.35% Indels: 0
DB: 2 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x T15969 (1-619)

QY 733 CACGCTTACGTGGGATCTCCCGCTCC 759
DB 223 HiscysleuAlaGlyIleSerArgSer 231

RESULT 2
A34402
peptidyl-dipeptidase A (EC 3.4.15.1) precursor, testicular - rabbit
N:Alternate names: angiotensin I-converting enzyme; dipeptidyl carboxypeptidase I; peptidyl
C:Species: *Oryctolagus cuniculus* (domestic rabbit)
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A34402; A60724; A36232; C18700
R:Kumar, R.S.; Kusari, J.; Roy, S.N.; Soifer, R.L.; Sen, G.C.
J. Biol. Chem. 264, 16754-16759, 1989
A:Title: Structure of testicular angiotensin-converting enzyme. A segmental mosaic isozy
A:Reference number: A34402; MUID:89380303; PMID:2550457
A:Accession: A34402
A:Molecule type: mRNA
A:Residues: 1-737 <KUM>
A:Cross-references: GB:J05041; NID:G164744; PIDN:AAA11153.1; PID:G164745
R:Sen, G.C.; Thekkumkara, T.U.; Kumar, R.S.
J. Cardiovasc. Pharmacol. 16(Suppl.4), S14-S18, 1990
A:Title: Angiotensin-converting enzyme: structural relationship of the testicular and the
A:Reference number: A60724; MUID:91155372; PMID:1705622
A:Accession: A60724
A:Status: translation not shown
A:Molecule type: mRNA
A:Residues: 73-173 <SEN>
A:Note: Identical sequences were obtained for mRNAs from lung and testes
R:Chen, Y.N.P.; Riordan, J.F.
Biochemistry 29, 10493-10498, 1990
A:Title: Identification of essential tyrosine and lysine residues in angiotensin convert
A:Reference number: A36232; MUID:91104959; PMID:2176870
A:Accession: A36232
A:Molecule type: protein
A:Residues: 154-160;235-242 <CHE>
R:Iwata, K.; Lai, C.Y.; El-Doray, H.A.; Soffer, R.L.
Biochem. Biophys. Res. Commun. 107, 1097-1103, 1982
A:Title: The NHD- and COOH-terminal sequences of the angiotensin-converting enzyme isozym
A:Reference number: A90107; MUID:83048249; PMID:6291514
A:Accession: C18700
A:Molecule type: protein
A:Residues: 33-35, 'SN', 38-39, 'SS', 'FAEL', 737 <IMA>
A:Note: Several of the amino acids in reported are tentative
C:Comment: The pulmonary and testicular isoforms of this enzyme differ substantially in
C:Superfamily: mammalian peptidyl-dipeptidase A
C:Keywords: alternative splicing; peptidyl-dipeptide hydrolase; testis; transmembrane pro

Alignment Scores:
Pred. No.: 4.63 Length: 737
Score: 9.00 Matches: 9

Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.40% Indels: 0
DB: 1 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x A34402 (1-737)

QY 1334 CATAGCTTGTGTCCTCCATCCAGACTG 1308
DB 22 H1SerLeuValProSerArgVal 30

RESULT 3
A97750
DNA mismatch repair protein Muts [imported] - *Rickettsia conorii* (strain Malish 7)
C:Species: *Rickettsia conorii*
C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001
C:Accession: A97750
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro
Science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in *Rickettsia conorii* and *Rickettsia prowazekii*.
A:Reference number: A97700; MUID:21442074; PMID:11557893
A:Accession: A97750
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-890 <KUR>
A:Cross-references: GB:AE006914; PIDN:AAL02939.1; PID:G15619468; GSPDB:GN00173
C:Genetics:
A:Gene: mutS
C:Superfamily: DNA mismatch repair protein muts

Alignment Scores:
Pred. No.: 4.5 Length: 890
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.40% Indels: 0
DB: 2 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x A97750 (1-890)

QY 1406 ATGCTGCTTCCTCTTATCAGACTG 1380
DB 572 MetLeuAlaSerSerLeuSerGlyLeu 580

RESULT 4
1B8G
hydroxymethylbilane synthase (EC 4.3.1.8) precursor - *Euglena gracilis*
N:Alternate names: porphobilinogen deaminase; pre-uroporphyrinogen synthase
C:Species: *Euglena gracilis*
C>Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 18-Jun-1999
R:Sharif, A.L.; Smuth, A.G.; Abell, C.
Eur. J. Biochem. 184, 353-359, 1989
A:Title: Isolation and characterisation of a cDNA clone for a chlorophyll synthetase enzy
nthesised with a very long transit peptide in *Euglena*.
A:Reference number: S06109; MUID:90005485; PMID:2477247
A:Accession: S06109
A:Molecule type: mRNA
A:Residues: 1-480 <SHA>
A:Cross-references: GB:X15743; NID:G18411; PIDN:CAA33759.1; PID:G18412
A:Note: part of this sequence, including the amino end of the mature protein, was confir
A:Note: 175-Ala was also found
C:Function:
A:Description: catalyzes the stepwise polymerization of four molecules of porphobilinoge
A:Pathway: porphyrin biosynthesis
A:Note: acting with uroporphyrinogen-III synthase (co-synthase), which cyclizes hydroxyme
C:Superfamily: hydroxymethylbilane synthase
C:Keywords: ammonia-lyase; carbon-nitrogen lyase; chlorophyll biosynthesis; chloroplast;
F:1-139/Domain: transit peptide (chloroplast) #status predicted <TMP>
F:140-480/Product: hydroxymethylbilane synthase #status experimental <MAT>
F:395/Modified site: dipyrrolyl-methanemethyl (Ys) (covalent) #status predicted

Alignment Scores:
Pred. No.: 4.63 Length: 737
Score: 9.00 Matches: 9

Pred. No.: 52.3 Length: 480
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 1 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x IREG (1-480)

QY 1722 CTACGAGGACGTCGACTTCTC 1745

DB 279 LeuArgArgGlnGlnSerLeuVal 286

RESULT 5

G86332

Protein 1 protein - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #ext_change 09-Nov-2001

C/Accession: G86332

R/Theologas, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

aneen, N.E.; Hughes, B.; Hutzar, L.

Nature 408, 816-820, 2000

A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maitl, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,

Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A/Reference number: A86141; MUID:21016719; PMID:11130712

A/Accession: G86332

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-515 <STO>

A/Cross-references: GB:AE005172; NID:G10086502; PIDN:AA012562.1; GSPDB:GN00141

C/Genetics:

A/Map position: 1

C/Superfamily: Arabidopsis membrane-anchored cellulase KOR

Alignment Scores:

Pred. No.: 51.8 Length: 515
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 2 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x G86332 (1-515)

QY 374 GTGGCTTGTGCTGCTCTCTCGTT 397

DB 499 ValGlyLeuLeuSerLeuVal 506

RESULT 6

E96786

protein F10A5.13 [imported] - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #ext_change 23-Mar-2001

C/Accession: E96786

R/Theologas, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

aneen, N.E.; Hughes, B.; Hutzar, L.

Nature 408, 816-820, 2000

A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maitl, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,

Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A/Reference number: A86141; MUID:21016719; PMID:11130712

A/Accession: E96786

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-525 <STO>
A/Cross-references: GB:AE005173; NID:G9369363; PIDN:AA07112.1; GSPDB:GN00141
C/Genetics:
A/Genes: F10A5.13
A/Map position: 1
C/Superfamily: Arabidopsis membrane-anchored cellulase KOR

Alignment Scores:

Pred. No.: 51.6 Length: 525
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 2 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x E96786 (1-525)

QY 374 GTGGCTTGTGCTGCTCTCTCGTT 397

DB 508 ValGlyLeuLeuSerLeuVal 515

RESULT 7

P5XR10

outer capsid protein VP5 - bluetongue virus (serotype 10, American isolate)

C/Species: bluetongue virus

C/Date: 28-Dec-1987 #sequence_revision 28-Dec-1987 #ext_change 16-Jun-2000

C/Accession: A25419; S10538

R/Purdy, M.A.; Ritzer, G.D.; Roy, P.

J. Gen. Virol. 67, 957-962, 1986

A/Title: Nucleotide sequence of cDNA clones encoding the outer capsid protein, VP5, of

A/Reference number: A25419; MUID:86198636; PMID:3009695

A/Accession: A25419

A/Molecule type: genomic RNA

A/Residues: 1-526 <PUR>

A/Cross-references: GB:D12532; GB:D01183; NID:G221078; PIDN:BAA02095.1; PID:G221079

R/Roy, P.; Marshall, J.J.A.; French, T.J.

Curr. Top. Microbiol. Immunol. 162, 43-87, 1990

A/Title: Structure of the bluetongue virus genome and its encoded proteins.

A/Reference number: S10534; MUID:90345726; PMID:2166548

A/Accession: S10538

A/Status: preliminary

A/Molecule type: genomic RNA

A/Residues: 1-526 <ROY>

C/Genetics:

A/Map position: segment 5

C/Superfamily: bluetongue virus outer capsid protein VP5

Alignment Scores:

Pred. No.: 51.6 Length: 526
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0
DB: 1 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x P5XR10 (1-526)

QY 98 CTATCATATGACGACCTTTTCC 75

DB 435 LeuSerLeuSerLeuSerLeuSer 442

RESULT 8

F95325

probable arylsulphatase [imported] - Sinorhizobium meliloti (strain 1021) megaplasmid pS1

C/Species: Sinorhizobium meliloti

C/Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #ext_change 30-Sep-2001

C/Accession: F95325

R/Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Aboja, A.P.; Barloy-Hubler, F.; Bow

, Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.C

Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001

A/Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti

A/Reference number: A95262; MUID:21396509; PMID:11481432

A:Accession: F95325
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-545 <KOR>
A:Cross-references: GB:AE006469; PIDN:AAK65168.1; PID:g14523612; GSPDB:GND00165
A:Experimental source: strain 1021, megaplasmid pSvMA
R:Gallibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abolaj, P.; Ampe, F.; Barloy-Hubler, P.; Chait, P.; Cowie, A.; Davies, R.W.; Dreano, S.; Federpiet, N.A.; Fisher, R.F.; L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure, M.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Weiler, D.H.; Wong, K.; Yeh, K.
A>Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
A:Reference number: A96039; PMID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: Sma0943
A:Genome: plasmid

Alignment Scores:
Pred. No.: 51.3 Length: 545
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0
DB: 2 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x F95325 (1-545)

QY 1593 CGTAGGTGCTGCTGCTGCTGCTGCA 1570
|||||
Db 15 ArgGlValIeuLeuLagIgly 22

RESULT 9
VGNZ11
cell fusion glycoprotein precursor - Sendai virus (strain C39)
N:Contains: fusion glycoprotein F1; fusion glycoprotein F2
C:Species: Sendai virus
C>Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 16-Jul-1999
C:Accession: A31287; S11284
R:Merison, J.R.; Hull, R.A.; Estes, M.K.; Kasel, J.A.
Virology 167, 97-105, 1988
A>Title: Molecular cloning and sequence determination of the fusion protein gene of human
A:Reference number: A31287; PMID:89045674; PMID:2847427
A:Accession: A31287
A:Molecule type: mRNA
A:Residues: 1-555 <MER>
A:Cross-references: GB:M22347; NID:g332579; PIDN:AAA46800.1; PID:g332580
A:Experimental source: strain C39
R:Richardson, C.D.; Scheid, A.; Choplin, P.W.
Virology 105, 205-222, 1980
A>Title: Specific inhibition of paramyxovirus and myxovirus replication by oligopeptides
A:Reference number: S11266; PMID:81016739; PMID:7414950
A:Accession: S11284
A:Molecule type: protein
A:Residues: 113-132 <RIG>
C:Genetics:
A:Gene: F
C:Superfamily: parainfluenza virus cell fusion protein
C:Keywords: glycoprotein; membrane fusion; transmembrane protein
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-112/Product: cell fusion glycoprotein F2 #status predicted <FR2>
F:113-555/Product: cell fusion glycoprotein F1 #status predicted <F1>
F:113-138/Domain: transmembrane #status predicted <TM>
F:148-519/Domain: transmembrane #status predicted <TM2>
F:100,241,529,552/Binding site: carbohydrate (Asn) (covalent) #status predicted

Alignment Scores:
Pred. No.: 51.2 Length: 555
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0

DB: 1 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x VGNZ11 (1-555)

QY 1193 ACCTATTGCTGCTTCCAGCCTG 1170
|||||
Db 13 SerLeuIeuSerSerSerIeu 20

RESULT 10
A35459
glucose oxidase (EC 1.1.3.4) precursor [validated] - Aspergillus niger
C:Species: Aspergillus niger
C>Date: 31-Aug-1990 #sequence_revision 09-May-1997 #text_change 20-Oct-2000
C:Accession: A35459; S05668; S14129
R:Frederick, K.R.; Tung, J.; Emerick, R.S.; Mastarz, F.R.; Chamberlain, S.H.; Vasavada, J. Biol. Chem. 265, 3793-3802, 1990
A>Title: Glucose oxidase from Aspergillus niger. Cloning, gene sequence, secretion from
A:Reference number: A35459; PMID:90154060; PMID:2406261
A:Accession: A35459
A:Molecule type: DNA
A:Residues: 1-605 <FR>
A:Cross-references: GB:J05242; NID:g166510; PIDN:AAA32695.1; PID:g166511
R:Krichbaum, M.; Hellmann, H.J.; Wientjes, F.J.; Hahn, W.; Jany, K.D.; Gassen, H.G.; Sh
FEBS Lett. 255, 63-66, 1989
A>Title: Cloning and DNA sequence analysis of the glucose oxidase gene from Aspergillus
A:Reference number: S05668; PMID:90005974; PMID:2792372
A:Accession: S05668
A:Molecule type: DNA
A:Residues: 1-605 <KR1>
A:Cross-references: EMBL:X16061; NID:g2356; PIDN:CAA34197.1; PID:g2357
A>Note: part of this sequence, including the amino end of the mature protein, was confir
R:Whittington, H.; Kerry-Williams, S.; Bidgood, K.; Dodsworth, N.; Peberdy, J.; Dobson,
Curr. Genet. 18, 531-536, 1990
A>Title: Expression of the Aspergillus niger glucose oxidase gene in A. niger, A. nidula
A:Reference number: S14129; PMID:91168301; PMID:2076553
A:Accession: S14129
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-32 <WH1>
A:Cross-references: GB:X56443; NID:g2354; PIDN:CAA39826.1; PID:g2355
C:Genetics:
A:Gene: gox
C:Superfamily: alcohol oxidase
C:Keywords: FAD; flavoprotein; oxidoreductase
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-605/Product: glucose oxidase #status experimental <MAT>
F:43-72/Region: beta-alpha-beta FAD nucleotide-binding fold

Alignment Scores:
Pred. No.: 50.5 Length: 605
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 1 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x A35459 (1-605)

QY 386 AGTCTCTGCTGCTTCCCTGACC 409
|||||
Db 7 SerSerLeuValValSerLeuA 14

RESULT 11
Q08E02
U35 protein - human cytomegalovirus (strain AD169)
C:Species: human cytomegalovirus, human herpesvirus 5
A>Note: host Homo sapiens (man)
C>Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1999
C:Accession: S09798
R:Chen, M.S.; Bankier, A.T.; Beck, S.; Bohni, R.; Brown, C.M.; Cerny, R.; Hornsneil, T.;
M.; Barrell, B.G.
Curr. Top. Microbiol. Immunol. 154, 125-169, 1990
A>Title: Analysis of the protein-coding content of the sequence of human cytomegalovirus

A:Reference number: S09749; MWID:90269039; PMID:2161319
A:Accession: S09798
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-640 <CHE>
A:Cross-references: EMBL:X17403; NID:G59591; PIDN:CAA5394.1; PID:G59640
A:Note: possible protein-coding frames are given
A:Note: the DNA sequence was submitted to EMBL, December 1989, in computer-readable form
C:Superfamily: cytomegalovirus UL35 protein

Alignment Scores:
Pred. No.: 50 Length: 640
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0
DB: 1 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x Q0BEU2 (1-640)

QY 1190 TTATTGCTGCTTCGAGCTGCT 1167

DB 547 LeuLeuLeuSerSerSerLeuSer 554

RESULT 12

F84524
En/Spm-like transposon protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C:Accession: F84524

R:Lin, X.; Kaul, S.; Rounstey, S.D.; Shea, T.P.; Benton, M.I.; Town, C.D.; Fujii, C.Y.;

M. Koo, H.; Mofatt, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.

Neus, D.; Nierman, W.C.; White, O.; Eisen, J.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MWID:20083487; PMID:10617197

A:Accession: F84524

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-672 <STO>

A:Cross-references: GB:AE002093; NID:G4115362; PIDN:AAD03364.1; GSPDB:GND0139

C:Genetics:

A:Gene: AT2G15070

A:Map position: 2

Alignment Scores:

Pred. No.: 49.7 Length: 672

Score: 8.00 Matches: 8

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 1.24% Indels: 0

DB: 2 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x F84524 (1-672)

QY 1094 ACCCTGGGCAACGCTGGCGGATGC 1071

DB 347 ThrLeuGlyThrLeuAlaGlyCys 354

RESULT 13

T22333
histidine ammoni-lyase (EC 4.3.1.3) - Caenorhabditis elegans

N:Alternate names: histidase; protein F47B10.2

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000

C:Accession: T22333

R:Harris, B.

submitted to the EMBL Data Library, November 1995

A:Reference number: Z19550

A:Accession: T22333

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-677 <WTL>

A:Cross-references: EMBL:Z68004; PIDN:CAA91982.1; GSPDB:GND0028; CESP:F47B10.2

A:Experimental source: clone F47B10

C:Genetics:

A:Gene: CESP:F47B10.2

A:Map position: X

A:Insertions: 24/3; 77/1; 111/2; 178/1; 255/1; 335/1; 399/1; 523/1; 544/3; 604/2

C:Superfamily: histidine ammoni-lyase

C:Keywords: ammoni-lyase; carbon-nitrogen lyase

F;269-271/Cross-link: 5-midazoliumone (Cys-Gly) #status predicted

F;270/Modified site: dehydroalanine (Ser) #status predicted

Alignment Scores:
Pred. No.: 49.6 Length: 677
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 2 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x T22333 (1-677)

QY 137 AAGCATTATATCAACTGCTCA 160

DB 584 LysProLeuIleSerThrAlaPro 591

RESULT 14

T52120
acyl-CoA oxidase (EC 1.3.3.6), peroxisomal [validated] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 03-Nov-2000

C:Accession: T52120

R:Hooks, M.A.; Kellas, F.; Graham, I.A.

Plant J. 20, 1-13, 1999

A:Title: Long-chain acyl-CoA oxidases of Arabidopsis.

A:Reference number: Z25966; MWID:20040045; PMID:10571860

A:Accession: T52120

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-692 <HOO>

A:Cross-references: EMBL:AF057043; PIDN:AAC13497.1

A:Experimental source: cultivar Columbia; seedling hypocotyl; 3 days old

C:Genetics:

A:Gene: ACX2

C:Function:

A:Description: EC 1.3.3.6 [validated, MWID:20040045]

A:Pathway: fatty acid beta-oxidation

A:Note: active with long-chain acyl-CoAs; maximal activity with C18-CoA

C:Superfamily: acyl-CoA oxidase

C:Keywords: fatty acid beta-oxidation; oxidoreductase; peroxisome

Alignment Scores:
Pred. No.: 49.4 Length: 692
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 2 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x T52120 (1-692)

QY 596 AGCGTACCTTATCCCGAGTCTC 619

DB 34 SerLeuThrLeuSerProSerLeu 41

RESULT 15

AB2353
hypothetical protein a114378 [imported] - Nostoc sp. (strain PCC 7120)

C:Species: Nostoc sp. PCC 7120

A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120

C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002

C:Accession: AB2353

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kunitz, T.; Sasamoto, S.; Watanabe, A.; Iriyuch;

Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, T.

DNA Res. 8, 205-213, 2001
A/Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Ana*
A/Reference number: AB1807; MUID:21595285; PMID:11759840
A/Accession: AB2353
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-727 <KUR>
A/Cross-references: GB:BA000019; PIDN:BA876077.1; PID:G17133514; GSPDB:GN00179
A/Experimental source: strain FCC 7120
C/Genetics:
A/Gene: all4378

Alignment Scores:
Pred. No.: 49.1 Length: 727
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 2 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x AB2353 (1-727)

OY 1081 AGCGTGGCAGCGCGCCAGCGTG 1104
Db 116 SerValProSerValProSerVal 123

RESULT 16
H90135
cell division cycle protein 48 homolog [imported] - *Guillardia theta* nucleomorph
C/Species: nucleomorph *Guillardia theta*
A/Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont
C/Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 15-Jun-2001
C/Accession: H90135
R/Douglas, S.; Zanner, S.; Franholcz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Re
Nature 410, 1091-1096, 2001
A/Title: The highly reduced genome of an enslaved algal nucleus.
A/Reference number: A99082; MUID:11323671; PMID:11323671
A/Status: preliminary
A/Accession: H90135
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-752 <DOU>
A/Cross-references: GB:AF083031; NID:G13794396; PIDN:AK39773.1; GSPDB:GN00152
C/Genetics:
A/Gene: cdc48
A/Map position: 3
A/Genome: nucleomorph
C/Superfamily: transitional endoplasmic reticulum ATPase, FtsH/SECT8/CDC48-type ATP-bind
C/Keywords: nucleomorph

Alignment Scores:
Pred. No.: 48.8 Length: 752
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0
DB: 2 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x H90135 (1-752)

OY 771 AGCGATGCGGCGGAGCGGAGAT 748
Db 570 SerAspGlyGlyAlaGlyAsp 577

RESULT 17
AE2688
N+/H+ antiporter mha [imported] - *Agrobacterium tumefaciens* (strain C58, Dupont)
C/Species: *Agrobacterium tumefaciens*
C/Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C/Accession: AE2688
R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, T
erage, G.; Gillet, W.; Grant, C.; Guenhner, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001

A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A/Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.
A/Reference number: AB2577; MUID:21608550; PMID:11743193
A/Accession: AE2688
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-789 <KUR>
A/Cross-references: GB:AE008688; PIDN:AAL41923.1; PID:G17739289; GSPDB:GN00186
A/Experimental source: strain C58 (Dupont)
C/Genetics:
A/Gene: mha
A/Map position: circular chromosome

Alignment Scores:
Pred. No.: 48.4 Length: 789
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0
DB: 2 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x AE2688 (1-789)

OY 1869 CTTTCAAGGGGCTCTTCATG 1846
Db 340 LeuPheLysGlyAlaLeuPheMet 347

RESULT 18
H97469
probable NADH dehydrogenase (VCA0157) [imported] - *Agrobacterium tumefaciens* (strain C58
C/Species: *Agrobacterium tumefaciens*
C/Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 18-Nov-2002
C/Accession: H97469
R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.
Science 294, 2323-2328, 2001
A/Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tun*
A/Reference number: A97359; MUID:21608551; PMID:11743194
A/Accession: H97469
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-789 <KUR>
A/Cross-references: GB:AE007869; PIDN:AK86713.1; PID:G15155503; GSPDB:GN00169
C/Genetics:
A/Gene: AGR_C_1658
A/Map position: circular chromosome

Alignment Scores:
Pred. No.: 48.4 Length: 789
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0
DB: 2 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x H97469 (1-789)

OY 1869 CTTTCAAGGGGCTCTTCATG 1846
Db 340 LeuPheLysGlyAlaLeuPheMet 347

RESULT 19
A27211
vira protein - *Agrobacterium tumefaciens* plasmid pTiAg162
C/Species: *Agrobacterium tumefaciens*
C/Date: 07-Mar-1988 #sequence_revision 07-Mar-1988 #text_change 08-Oct-1999
C/Accession: A27211
R/Leroux, B.; Yanofsky, M.F.; Winans, S.C.; Ward, J.E.; Ziegler, S.F.; Neeter, E.W.
EMBO J. 6, 849-856, 1987
A/Title: Characterization of the vira locus of *Agrobacterium tumefaciens*: a transcriptio
A/Reference number: A91082; MUID:87246518; PMID:3595559
A/Accession: A27211

A:Molecule type: DNA
A:Residues: 1-835 <EB>
A:Cross-references: GB:X05241; NID:g39191; PIDN:CAA28868.1; PID:g39192
C:Genetics:
A:Gene: vira
A:Genome: Plasmid
C:Keywords: autophosphorylation; host range; phosphoprotein; transcription regulation; v

Alignment Scores:
Pred. No.: 48 Length: 835
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0
DB: Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x A27211 (1-835)

QY 545 ATCAGCTCTGTTGAGACATCT 522
DB 148 lleserSerleuValgthrSer 155

RESULT 20
T12386
hypothetical protein F29D1.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C:Accession: T12386; T12549
R:McMurray, A.
submitted to the EMBL Data Library, August 1996
A:Reference number: Z19415
A:Accession: T21386
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1153 <WIL>
A:Cross-references: EMBL:Z78419; PIDN:CAB01707.1; GSPDB:GN00019; CESP:F29D11.2
A:Experimental source: clone F26A3
R:Wilkinson, J.
submitted to the EMBL Data Library, June 1996
A:Reference number: Z19439
A:Accession: T21549
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1153 <WIL>
A:Cross-references: EMBL:Z73907; PIDN:CAA98126.1; GSPDB:GN00019; CESP:F29D11.2
A:Experimental source: clone F29D11
C:Genetics:
A:Gene: CESP:F29D11.2
A:Map position: 1
A:Introns: 83/2; 117/3; 186/3; 243/3; 278/1; 353/1; 493/2; 543/2; 685/3; 884/2; 944/3; 1
C:Superfamily: Caenorhabditis elegans hypothetical protein F29D11.2

Alignment Scores:
Pred. No.: 45.7 Length: 1153
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0
DB: Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x T12386 (1-1153)

QY 891 CTCATAGTCCAGAGTTGGCCGAG 868
DB 615 leuileValGlnleuValGln 622

RESULT 21
T18734
hypothetical protein B0391.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18734
R:Gardner, A.

submitted to the EMBL Data Library, November 1996
A:Reference number: Z19012
A:Accession: T18734
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1286 <WIL>
A:Cross-references: EMBL:Z81454; PIDN:CAB03805.1; GSPDB:GN00023; CESP:B0391.6
A:Experimental source: clone B0391
C:Genetics:
A:Gene: CESP:B0391.6
A:Map position: 5
A:Introns: 64/2; 277/3; 350/2; 427/2; 494/2; 522/2; 749/3; 806/1; 872/2; 898/2; 1111/3;

Alignment Scores:
Pred. No.: 44.9 Length: 1286
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x T18734 (1-1286)

QY 383 CTCAGTCTCGCTGTTCCCTG 406
DB 1254 leuSerSerleuValValSerleu 1261

RESULT 22
T15280
hypothetical protein R155.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15280
R:Geisel, C.; Wamsley, P.; Kramer, J.
submitted to the EMBL Data Library, May 1997
A:Description: The sequence of C. elegans cosmid R155.
A:Reference number: Z18321
A:Accession: T15280
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1308 <GEI>
A:Cross-references: EMBL:AF003390; NID:G2088866; PID:G2088869; PIDN:AAB54272.1; GSPDB:G
A:Experimental source: strain Bristol N2; clone R155
C:Genetics:
A:Gene: CESP:R155.2
A:Map position: 3
A:Introns: 13/1; 57/1; 860/2; 897/1; 970/2; 1116/2; 1174/1; 1225/3

Alignment Scores:
Pred. No.: 44.8 Length: 1308
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x T15280 (1-1308)

QY 745 GGGATCTCCGCTCCGACCATC 768
DB 1165 GlyileSerArgSerAlaThrIle 1172

RESULT 23
T14103
probable DNA-directed RNA polymerase (EC 2.7.7.6) I large chain - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Mar-2000
C:Accession: T14103
R:Hanman, R.D.; Hemple, W.; Rothblum, L.I.
submitted to the EMBL Data Library, September 1997
A:Reference number: Z17875
A:Accession: T14103
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA
A:Residues: 1-1716 <HMAN>
A:Cross-references: EMBL:AF025425; NID:g2739049; PID:g2739050; PIDN:AAB94601.1
A:Experimental source: Reuber hepatoma
A:Genetics:
A:Gene: RPAL
C:Superfamily: human DNA-directed RNA polymerase II largest chain
C:Keywords: nucleotidyltransferase

Alignment Scores:
Pred. No.: 42.9 Length: 1716
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 2 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x T14103 (1-1716)

QY 386 AGTTCCTCGTTGTTTCCTGGCC 409

Db 1537 SerSerLeuValValSerLeuAla 1544

RESULT 24

T13961
DNA-directed RNA polymerase (EC 2.7.7.6) I large chain - mouse
N:Contains: nonstructural protein NS1; nonstructural protein NS2; nonstructural protein
C:Species: Mus musculus (house mouse)
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Mar-2000
C:Accession: T13961
R:Seither, P.; Coy, J.F.; Pouska, A.; Grummt, I.
Mol. Gen. Genet. 255, 180-186, 1997
A>Title: Molecular cloning and characterization of the cDNA encoding the largest subunit
A:Reference number: Z17838; MUID:97379976; PMID:9236775
A:Accession: T13961
A:Status: preliminary; translated from GB/EMBL/DBEST
A:Molecule type: mRNA
A:Residues: 1-1717 <SET>
A:Cross-references: EMBL:AF000938; NID:g2330006; PID:g2330007; PIDN:AAB66718.1
A:Genetics:
A:Gene: RPAL
C:Superfamily: human DNA-directed RNA polymerase II largest chain
C:Keywords: nucleotidyltransferase

Alignment Scores:
Pred. No.: 42.9 Length: 1717
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 2 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x T13961 (1-1717)

QY 386 AGTTCCTCGTTGTTTCCTGGCC 409

Db 1538 SerSerLeuValValSerLeuAla 1545

RESULT 25

MMNV8
nonstructural polypeptide - Sindbis virus
N:Contains: nonstructural NS2; nonstructural protein NS1; nonstructural protein NS3; non
C:Species: Sindbis virus
C>Date: 19-Feb-1984 #sequence_revision 03-Aug-1984 #text_change 04-Oct-1996
C:Accession: A03917
R:Straus, E.G.; Rice, C.M.; Straus, J.H.
Virology 133, 92-110, 1984
A>Title: Complete nucleotide sequence of the genomic RNA of Sindbis virus.
A:Reference number: A94331; MUID:84148439; PMID:6322438
A:Accession: A03917
A:Molecule type: genomic RNA
A:Residues: 1-2512 <STR>
A:Experimental source: strain HRSP

A>Note: readthrough of the terminator UGA between codons UAC for 1896-Tyr and CUA for 1
C:Superfamily: Semliki Forest virus nonstructural protein
C:Keywords: polypeptide
F:1-540/Product: nonstructural protein NS1 #status predicted <NS1>
F:541-1347/Product: nonstructural protein NS2 #status predicted <NS2>
F:1348-1896/Product: nonstructural protein NS3 #status predicted <NS3>
F:1897-2512/Product: nonstructural protein NS4 #status predicted <NS4>

Alignment Scores:
Pred. No.: 40.5 Length: 2512
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 1 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x MMNV8 (1-2512)

QY 1550 CCAGCTTCCTTTGGCCCTTCCA 1573

Db 445 ProLaserPheSerAlaPhePro 452

RESULT 26

MMNV82
nonstructural polypeptide - Ockelbo virus (strain R48bvn 82-5)
N:Contains: nonstructural protein NS1; nonstructural protein NS2; nonstructural protein
C:Species: Ockelbo virus
C>Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 08-Apr-1994
C:Accession: A39991
R:Shirako, Y.; Niklasson, B.; Dalrymple, J.M.; Straus, E.G.; Straus, J.H.
Virology 182, 753-764, 1991
A>Title: Structure of the Ockelbo virus genome and its relationship to other Sindbis vi
A:Reference number: A39991; MUID:91220725; PMID:1673813
A:Accession: A39991
A:Molecule type: genomic RNA
A:Residues: 1-2514 <SHI>
A:Cross-references: GB:M69205
C:Superfamily: Semliki Forest virus nonstructural protein
C:Keywords: nonstructural protein; polypeptide
F:1-540/Product: nonstructural protein NS1 #status predicted <NS1>
F:541-1347/Product: nonstructural protein NS2 #status predicted <NS2>
F:1348-1898/Product: nonstructural protein NS3 #status predicted <NS3>
F:1899-2514/Product: nonstructural protein NS4 #status predicted <NS4>

Alignment Scores:
Pred. No.: 40.5 Length: 2514
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 1 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x MMNV82 (1-2514)

QY 1550 CCAGCTTCCTTTGGCCCTTCCA 1573

Db 445 ProLaserPheSerAlaPhePro 452

RESULT 27

T29285
hypothetical protein C34D4.14 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T29285
R:Du, Z.; Le, T.T.
submitted to the EMBL Data Library, May 1996
A:Description: The sequence of C. elegans cosmid C34D4.
A:Reference number: Z20600
A:Accession: T29285
A:Status: preliminary; translated from GB/EMBL/DBEST
A:Molecule type: DNA
A:Residues: 1-2761 <DUZ>
A:Cross-references: EMBL:U58755; PIDN:AAB00699.1; GSPDB:GN00022; CESP:C34D4.14

A:Experimental source: strain Bristol N2; clone C34D4
C:Genetics:
A:Gene: CBSP:C34D4.14
A:Map position: 4
A:Insertions: 120/1; 201/3; 365/3; 445/2; 509/1; 728/3; 856/3; 902/3; 940/3; 1027/3; 1508/3
Alignment Scores:
Pred. No.: 39.9 Length: 2761
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0
Gaps: 0
DB: 2
US-10-029-345A-108_COPY_532_2532 (1-2001) x T29285 (1-2761)
QY 1743 GTAACGGACAGCTGCTCGTAGAT 1720
|||||
Db 841 ValThrGlyThrAlaSerValAsp 848
RESULT 28
RBHUAP
adenomatous polyposis coli protein - human
N:Alternate names: polyposis coli locus protein DP2.5; tumor suppressor APC
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 21-Jul-2000
A:Accession: AJ37261; B39658; A49319; I54271
R:Kinzler, K.W.; Nishisho, M.C.; Su, L.K.; Vogelstein, B.; Bryan, T.M.; Levy, D.B.; Smith
chul, S.F.; Horii, A.; Ando, H.; Miyoshi, Y.; Miki, Y.; Nishisho, I.; Nakamura, Y.
Science 253, 661-665, 1991
A:Title: Identification of FAP locus genes from chromosome 5q21.
A:Reference number: AJ37261; MUID:91335210; PMID:1651562
A:Accession: AJ37261
A:Molecule type: mRNA
A:Residues: 1-2843 <KIN>
A:Cross-references: GB:M74088; NID:g182396; PIDN:AA03586.1; PID:g182397
R:Jostlyn, G.; Carlson, M.; Thliveris, A.; Albertsen, H.; Gelbert, L.; Samowitz, W.; Grod
arlington, J.; Mcherson, J.; Wasmuth, J.; Le Paslier, D.; Abderrahim, H.; Cohen, D.; Le
Cell 66, 601-613, 1991
A:Title: Identification of deletion mutations and three new genes at the familial polyp
A:Reference number: A39658; MUID:91330307; PMID:1678319
A:Accession: B39658
A:Molecule type: DNA
A:Residues: 1-183, 'L', 185-969, 'N', 971-1108, 'G', 1310-1324, 'SS', 1326, 'HSTLE', 1332-1354, 'P'
A:Cross-references: GB:M73548; NID:g190163; PIDN:AA60354.1; PID:g190164
R:Miki, Y.; Nishisho, I.; Horii, A.; Miyoshi, Y.; Utsunomiya, J.; Kinzler, K.W.; Vogelst
Cancer Res. 52, 643-645, 1992
A:Title: Disruption of the APC gene by a retrotransposon insertion of L1 sequence in a c
A:Reference number: A49328; MUID:92119623; PMID:1310068
A:Accession: A49328
A:Molecule type: DNA
A:Residues: 1506-1525 <MIX>
A:Cross-references: GB:578214; NID:g243541; PIDN:AA21145.1; PID:g243542
R:Note: sequence extracted from NCI backbone (NCBI:78214, NCBI:P:78218)
R:Sp1to, L.; Olschewski, S.; Groden, J.; Robertson, M.; Samowitz, W.; Jostlyn, G.; Gelber
Cell 75, 951-957, 1993
A:Title: Alleles of the APC gene: an attenuated form of familial polyposis.
A:Reference number: A49319; MUID:9407973; PMID:8252630
A:Accession: A49319
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 'G', 143-171, 'P', 173-179 <SP1>
A:Cross-references: GB:567787; NID:g461061; PIDN:AA013997.1; PID:g4261697
R:Lambertz, S.; Ballhausen, W.G.
Hum. Genet. 90, 650-652, 1993
A:Title: Identification of an alternative 5' untranslated region of the adenomatous poly
A:Reference number: I54271; MUID:93186137; PMID:8383094
A:Accession: I54271
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-4 <LAW>
A:Cross-references: GB:56365; NID:g26243; PIDN:AA01918.1; PID:g4262770
C:Genetics:

A:Gene: GDB:APC
A:Cross-references: GDB:119682; OMIM:175100
A:Map position: 5q21-5q22
A:Note: mutations of this gene can result in familial adenomatous polyposis or sporadic
C:Superfamily: adenomatous polyposis coli protein
C:Keywords: cancer; familial adenomatous polyposis; tumor suppressor
F:1-730/Domain: leucine-rich <NTD>
F:7-72/Region: coli #status predicted
F:185-227/Region: coli #status predicted
F:731-2832/Domain: serine-rich <CTD>
F:1131-1156/Region: acidic
F:1558-1577/Region: acidic
F:1866-1893/Region: highly charged
Alignment Scores:
Pred. No.: 39.7 Length: 2843
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0
Gaps: 0
DB: 1
US-10-029-345A-108_COPY_532_2532 (1-2001) x RBHUAP (1-2843)
QY 959 TTCTCCAGGTGACGAGCTTGAGT 936
|||||
Db 1271 PheSerArgCysSerSerLeuSer 1278
RESULT 29
149505
adenomatous polyposis coli protein - mouse
N:Alternate names: APC
C:Species: Mus musculus (house mouse)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 13-Aug-1999
A:Accession: I49505
R:Su, L.
Science 256, 668-670, 1992
A:Title: Multiple intestinal neoplasia caused by a mutation in the murine homolog of th
A:Reference number: I49505; MUID:92263101; PMID:1350108
A:Accession: I49505
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2845 <RES>
A:Cross-references: GB:M88127; NID:g191991; PIDN:AA05632.1; PID:g191992
C:Superfamily: adenomatous polyposis coli protein
Alignment Scores:
Pred. No.: 39.7 Length: 2845
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0
Gaps: 0
DB: 2
US-10-029-345A-108_COPY_532_2532 (1-2001) x I49505 (1-2845)
QY 959 TTCTCCAGGTGACGAGCTTGAGT 936
|||||
Db 1269 PheSerArgCysSerSerLeuSer 1276
RESULT 30
B97073
beta-glucosidase [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
A:Accession: B97073
R:Nozling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee
; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cl
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: B97073
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-473 <KUR>
A:Cross-references: GB:AF001437; PIDN:AAK79373.1; PID:G15024344; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Gene: CAC1405
C:Superfamily: Agrobacterium beta-glucosidase

Alignment Scores:
Pred. No.: 554 Length: 473
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.09% Indels: 0
DB: 2 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x B97073 (1-473)

QY 1791 ATGACTTGCTCCGCAAGT 1771
|||||
Db 422 lIeAspIeuValSerAlaSer 428

RESULT 31
D91075
6-phospho-beta-glucosidase [imported] - Escherichia coli (strain O157:H7, substrain R1MD
C/Species: Escherichia coli
C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C/Accession: D91075
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gasawara, N.; Yasunaga, T.; Kunara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: D91075
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-474 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA836995.1; PID:G13363043; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain R1MD 0509952
C:Gene: C
A:Gene: EC63572
C:Superfamily: Agrobacterium beta-glucosidase

Alignment Scores:
Pred. No.: 554 Length: 474
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.09% Indels: 0
DB: 2 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x D91075 (1-474)

QY 1791 ATGACTTGCTCCGCAAGT 1771
|||||
Db 422 lIeAspIeuValSerAlaSer 428

RESULT 32
C85920
6-phospho-beta-glucosidase, cryptic [imported] - Escherichia coli (strain O157:H7, subst
C/Species: Escherichia coli
C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C/Accession: C85920
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Iller, L.; Grobbeck, E.J.; Davis, N.W.; Lhm, A.; Dimlantia, E.; Potamouits, K.; Apodaca,
Nature 409, 529-533, 2001
A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: AB5480; MUID:21074935; PMID:11206551
A:Accession: C85920
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-474 <STO>
A:Cross-references: GB:AB005174; NID:g12517164; PIDN:AMG57823.1; GSPDB:GN00145; UWGP:Z40

A:Experimental source: strain O157:H7, substrain EDJ933
C:Gene: ascB
A:Superfamily: Agrobacterium beta-glucosidase

Alignment Scores:
Pred. No.: 554 Length: 474
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.09% Indels: 0
DB: 2 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x H65051 (1-474)

QY 1791 ATGACTTGCTCCGCAAGT 1771
|||||
Db 422 lIeAspIeuValSerAlaSer 428

RESULT 33
H65051
6-phospho-beta-glucosidase (BC 3.2.1.86) - Escherichia coli (strain K-12)
C/Species: Escherichia coli
C/Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C/Accession: H65051; C44070; G27553
R:Plattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A>Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: H65051
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-474 <BLAT>
A:Cross-references: GB:AF000355; GB:U00096; NID:G2367151; PIDN:AAK75758.1; PID:G1789070,
R:Hall, B.G.; Xu, L.
Mol. Biol. Evol. 9, 688-706, 1992
A>Title: Nucleotide sequence, function, activation, and evolution of the cryptic asc ope
A:Reference number: A44070; MUID:92334140; PMID:1630307
A:Accession: C44070
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-404, 'GT', 407-427, 'C', 429-454, 'HR', 457-474 <HAL>
A:Cross-references: GB:W33326; NID:G145385; PIDN:AAAI6430.1; PID:G145388
A:Experimental source: strain LP103
A>Note: sequence extracted from NCBI backbone (NCBI:109109, NCBI:P.109114)
C:Gene: ascB
A:Map position: 59 min
C:Superfamily: Agrobacterium beta-glucosidase
C:Keywords: glycosidase; hydrolase

Alignment Scores:
Pred. No.: 554 Length: 474
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.09% Indels: 0
DB: 2 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x H65051 (1-474)

QY 1791 ATGACTTGCTCCGCAAGT 1771
|||||
Db 422 lIeAspIeuValSerAlaSer 428

RESULT 34
S49886
probable membrane protein YII123w - Yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YI1877.06
C/Species: Saccharomyces cerevisiae
C/Date: 13-Jan-1995 #sequence_revision 20-Feb-1995 #text_change 11-Jan-2000

C/Accession: S49886
 R/Hamlyn, N.; Churcher, C.
 Submitted to the EMBL Data Library, November 1994
 A/Reference number: S49881
 A/Accession: S49886
 A/Molecule type: DNA
 A/Residues: 1-475 <KLE>
 A/Cross-references: GB:247047; EMBL:246833; NID:g603997; PID:g763223; MIPS:YLL123W
 C/Genetics:
 A/Gene: SGD:SIM1
 A/Cross-references: SGD:S0001385; MIPS:YLL123W
 A/Map position: 9L
 C/Superfamily: Saccharomyces NC3 protein
 C/Keywords: transmembrane protein
 F/82-98/Domain: transmembrane #status predicted <TM>

Alignment Scores:
 Pred. No.: 554 Length: 475
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.09% Indels: 0
 DB: 2 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x S49886 (1-475)

QY 1058 CCTGCTGCTCTGAGTAGCA 1038
 DB 57 ProAlaAlaSerGluValAla 63

RESULT 35
 B70143
 fibronectin/fibrinogen-binding protein homolog - Lyme disease spirochete
 C/Species: Borrelia burgdorferi (Lyme disease spirochete)
 C/Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
 C/Accession: B70143
 R/Faaser, C.M.; Caslens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
 son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugr,
 Nature 390, 580-586, 1997
 A/Authors: Smith, H.O.; Venter, J.C.
 A/Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
 A/Reference number: A70100; MUID:98065943; PMID:9403685
 A/Accession: B70143
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: DNA
 A/Residues: 1-475 <KLE>
 A/Cross-references: GB:AE001141; GB:AE000783; NID:g2688250; PIDN:AA06734.1; PID:g268825
 A/Experimental source: strain B31

Alignment Scores:
 Pred. No.: 554 Length: 475
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.05% Indels: 0
 DB: 2 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x B70143 (1-475)

QY 892 AAGAAGTTAAGACCGACT 912
 DB 469 LysLysLysLysSerGlnThr 475

RESULT 36
 H89773
 hypothetical protein SA0127 [imported] - Staphylococcus aureus (strain N315)
 C/Species: Staphylococcus aureus
 C/Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
 C/Accession: H89773
 R/Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogun
 ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
 C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001
 A/Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
 A/Reference number: A89758; MUID:21311952; PMID:11418146
 A/Accession: H89773
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-476 <KUR>
 A/Cross-references: GB:BA000018; PID:g13700048; PIDN:BA041347.1; GSPDB:GN00149
 A/Experimental source: strain N315
 C/Genetics:
 A/Gene: SA0127

Alignment Scores:
 Pred. No.: 554 Length: 476
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.05% Indels: 0
 DB: 2 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x H89773 (1-476)

QY 452 AGCCTTGCTTACCTGTTGCCA 472
 DB 211 SerLeuAlaTyrLeuLeuPro 217

RESULT 37
 E97073
 phospho-beta-glucosidase [imported] - Clostridium acetobutylicum
 C/Species: Clostridium acetobutylicum
 C/Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
 C/Accession: E97073
 R/Molling, U.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee
 J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
 J. Bacteriol. 183, 4823-4838, 2001
 A/Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium CJ
 A/Reference number: A96900; MUID:21359325; PMID:21359325
 A/Accession: E97073
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-477 <KUR>
 A/Cross-references: GB:AE001437; PIDN:AAK79376.1; PID:g15024347; GSPDB:GN00168
 A/Experimental source: Clostridium acetobutylicum ATCC824
 C/Genetics:
 A/Gene: CAC1408
 C/Superfamily: Agrobacterium beta-glucosidase

Alignment Scores:
 Pred. No.: 553 Length: 477
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.09% Indels: 0
 DB: 2 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x E97073 (1-477)

QY 1791 ATAGACTGTGTCGCGAAGT 1771
 DB 425 IleAspLeuValSerAlaSer 431

RESULT 38
 G83048
 probable two-component sensor PA4777 [imported] - Pseudomonas aeruginosa (strain PA01)
 C/Species: Pseudomonas aeruginosa
 C/Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C/Accession: G83048
 R/Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kass, A.; Lapidig, K.; Lim
 Nature 406, 959-964, 2000
 A/Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic path
 A/Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: G83048
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-477 <STO>
A:Cross-references: GB:AE004891; GB:AE004091; NID:g95951037; PIDN:AA608163.1; GSPDB:GN001
A:Experimental source: strain PMOI
C:Genetics:
A:Gene: PA4777

Alignment Scores:
Pred. No.: 553 Length: 477
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.09% Indels: 0
DB: 2 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x G83048 (1-477)

QY 1841 CTCGCGCGGAGTCAGCTCTG 1821
Db 90 LeuArgArgGlnSerAlaLeu 96

RESULT 39
C68807
beta-glucosidase (EC 3.2.1.21) [imported] - Lactococcus lactis subsp. lactis (strain IL1)
C:Species: Lactococcus lactis subsp. lactis
C>Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: C68807
R:Bojotin, A.; Wincker, P.; Manger, S.; Jallou, O.; Malarme, K.; Weissenbach, J.; Ehrlich
Genome Res. 11, 731-753, 2001
A>Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis ss
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: C68807
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-478 <STO>
A:Cross-references: GB:AE005176; PID:g12724451; PIDN:AAK05557.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: bgjH
C:Superfamily: Agrobacterium beta-glucosidase
C:Keywords: glycosidase; hydrolase

Alignment Scores:
Pred. No.: 553 Length: 478
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.09% Indels: 0
DB: 2 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x C68807 (1-478)

QY 1791 ATAGACTTGCTCCGCAAGT 1771
Db 427 IleAspLeuValSerAlaSer 433

RESULT 40
S17836
fatty-acyl-CoA reductase (EC 1.2.1.-) luxC - Photobacterium leiognathi
C:Species: Photobacterium leiognathi
C>Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 26-May-2000
C:Accession: S17836
R:Lee, C.Y.; Sziltnier, R.B.; Weighen, E.A.
Eur. J. Biochem. 201, 161-167, 1991
A>Title: The lux genes of the luminous bacterial symbiont, Photobacterium leiognathi, of
coll.
A:Reference number: S17836; MUID:92007870; PMID:1915359
A:Accession: S17836
A:Molecule type: DNA
A:Residues: 1-478 <LEE>
A:Cross-references: EMBL:M63594; NID:g150687; PIDN:AAA25616.1; PID:g150688

C:Genetics:
A:Gene: luxC
C:Superfamily: Photobacterium fatty-acyl-CoA reductase
C:Keywords: oxidoreductase

Alignment Scores:
Pred. No.: 553 Length: 478
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.05% Indels: 0
DB: 2 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x S17836 (1-478)

QY 355 TCTGTTACCTGCTTCAGGT 375
Db 145 SerValHisLeuLeuAlaGly 151

RESULT 41
F70573
hypothetical protein Fv0341 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: F70573
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
Rajandream, M.A.; Rogers, R.; Devlin, K.; Felkwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: F70573
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-479 <COL>
A:Cross-references: GB:Z95324; GB:AL123456; NID:g3261760; PIDN:CAB08573.1; PID:g2094820
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Fv0341
C:Superfamily: Phaeolus glycine-rich cell wall protein 1.8

Alignment Scores:
Pred. No.: 553 Length: 479
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.09% Indels: 0
DB: 2 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x F70573 (1-479)

QY 1491 GCGGTGCGCACTGCTGCTGCT 1471
Db 342 GlycylAlaThrAlaAlaGly 348

RESULT 42
T47415
transporter-like protein - Arabidopsis thaliana
N:Alternate names: protein T28A8.80
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T47415
R:Purnelle, B.; Boutry, M.; Goffeau, A.; Mewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X
submitted to the Protein Sequence Database, March 2000
A:Reference number: Z24466
A:Accession: T47415
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-479 <PUR>
A:Cross-references: EMBL:AL162691
A:Experimental source: cultivar Columbia; BAC clone T28A8
C:Genetics:

A:Map position: 3
A:Introns: 48/1: 77/1; 111/2; 145/3; 163/3; 184/3; 201/2; 226/3; 291/3; 321/1; 349/3; 38
A:Note: T28A8.80

Alignment Scores:

Score:	553	Length:	479
Pred. No.:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.05%	Indels:	0
DB:	2	Gaps:	0

US-10-029-345A-108_COPY_532_2532 (1-2001) x T47415 (1-479)

QY 1558 CTTTGGCCTTCACACAC 1578
|||||
DB 117 Leupheg1yenseRhrser 123

RESULT 43

H86411
Protein PK23.12 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: H86411
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
amen, N.R.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marzall,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,
Ker, M.; Wu, D.; Yu, G.; Fraser, C.W.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: H86411
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-480 <STO>
A:Cross-references: GB:AE005172; NID:G1074863; PIDN:AAF24555.2; GSPDB:GN00141
C:Gene(s):
A:Gene: PK23.12
A:Map position: 1

Alignment Scores:

Pred. No.:	553	Length:	480
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.09%	Indels:	0
DB:	2	Gaps:	0

US-10-029-345A-108_COPY_532_2532 (1-2001) x H86411 (1-480)

QY 57 AGGCACCACTCTCAGTAC 37
|||||
DB 15 SerHidg1ProleuserAen 21

RESULT 44

H69593
6-phospho-beta-glucosidase (EC 3.2.1.86) bgIH [similarity] - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 21-Jul-2000
C:Accession: H69593; I40407; T47098; S47175
R:Kuner, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Beret,
C.; Bron, S.; Broutlet, S.; Bruschi, C.V.; Caldwell, B.; Capuani, G.; Carter, N.M.; Chd
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallier
lech, J.; Harwood, C.R.; Henauf, A.; Hilbert, H.; Holappel, S.; Hosono, S.; Hullo, M.F.
Koeter, P.; Koningsreide, G.; Krogh, K.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A:Authors: Lamber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelli

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scanlon
A:Authors: Schlach, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowske, A.; Sero
akeuchi, M.; Tanakoshi, A.; Tanaka, T.; Tappera, P.; Tognoni, A.; Tosato, V.; Uchiyama
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida,
A:Authors: Yoshikawa, H.F.; Zumestein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033; PMID:9384377
A:Accession: H69593
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-481 <KUN>
A:Cross-references: GB:Z99124; GB:AL009126; NID:G2636442; PIDN:CA815962.1; PID:G2636472
A:Experimental source: strain 168
R:Le Cog, D.; Lindner, C.; Kruger, S.; Steinmetz, M.; Stulke, J.
U. Bacteriol. 177, 1527-1535, 1995
A:Title: New beta-glucosidase (bgII) genes in Bacillus subtilis: the bgII gene product has
A:Reference number: I40404; MUID:95189730; PMID:7883710
A:Accession: I40407
A>Status: preliminary; translated from GB/EMBL/DBSJ
A:Molecule type: DNA
A:Residues: 13-413, R', 415-481 <LEC>
A:Cross-references: EMBL:Z34526; NID:G505573; PIDN:CAA84287.1; PID:G505577
R:Yoshida, K.; Sano, H.; Seki, S.; Oda, M.; Fujimura, M.; Fujita, Y.
Microbiology 141, 337-343, 1995
A:Title: Cloning and sequencing of a 29 kb region of the Bacillus subtilis genome conta
A:Reference number: Z24350; MUID:95219088; PMID:7704263
A:Accession: I47098
A>Status: preliminary; translated from GB/EMBL/DBSJ
A:Molecule type: DNA
A:Residues: 1-481 <ROS>
A:Cross-references: EMBL:D31856; NID:G603765; PIDN:BAA06653.1; PID:G603779
A:Experimental source: strain BGSC1A1
C:Gene(s):
A:Gene: bgIIH; N17D
C:Superfamily: Agrobacterium beta-glucosidase
C:Keywords: glycosidase; hydrolase

Alignment Scores:

Pred. No.:	553	Length:	481
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.09%	Indels:	0
DB:	2	Gaps:	0

US-10-029-345A-108_COPY_532_2532 (1-2001) x H69593 (1-481)

QY 1791 ATGACTGTCTCGCAGT 1771
|||||
DB 430 Ileapleuvalseralaser 436

RESULT 45

AB1421
beta-glucosidase homolog lmo2771 [imported] - Listeria monocytogenes (strain EGD-e)
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C:Accession: AB1421
R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Anand, A.; Baquero, F.; Berche, P.; Blocke,
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duesuguet, O.; Entian, K.D.; Feihl, H
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Krefel, U.; Kuhn, M.; Kunst, F.; Kurapkut, G.; Madueno, E.; Maltournam, A.; M
Ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlund,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AB1421
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-482 <GLA>
A:Cross-references: GB:NC_003210; PIDN:CAD00984.1; PID:G16412271; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Gene(s):
A:Gene: lmo2771

C:/Superfamily: Agrobacterium beta-glucosidase

Alignment Scores:

Pred. No.:	552	Length:	482
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.09%	Indels:	0
DB:	2	Gaps:	0

US-10-029-345A-108_COPY_532_2532 (1-2001) x AB1421 (1-482)

OY 1791 ATGAGCTGTCTCCGCAAGT 1771

Db 428 lIeapleuValSerAlaSer 434

RESULT 46

A31521

Orphan receptor TR2, splice form TR2-5 - human

C/Species: Homo sapiens (man)

C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000

C/Accession: A31521

R/Chang, C.; Kokontis, J.

Biochem. Biophys. Res. Commun. 155, 971-977, 1988

A/Title: Identification of a new member of the steroid receptor super-family by cloning

A/Reference number: A31521; MUID:8833993; PMID:3421977

A/Accession: A31521

A/Molecule type: mRNA

A/Residues: 1-483 <CNA>

A/Cross-references: GB:M21985; NID:g338485; PIDN:AAA3650.1; PID:g338486

C/Genetics:

A/Gene: GDB:TR2-11

A/Cross-references: GDB:5875387

C/Superfamily: orphan receptor TR2; erba transforming protein homology

C/Keywords: alternative splicing; DNA binding; nucleus; receptor; transcription regulati

F/113-133/Region: erba transforming protein homology #status atypical <ERBA>

F/149-168/Region: zinc finger CCCC motif

Alignment Scores:

Pred. No.:	552	Length:	483
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.09%	Indels:	0
DB:	1	Gaps:	0

US-10-029-345A-108_COPY_532_2532 (1-2001) x A31521 (1-483)

OY 487 GGCTGGCCCAATGTTGGCA 467

Db 437 gLyLeuAlaGlnCysTrpGln 443

RESULT 47

F87700

rfaE protein [imported] - Caulobacter crescentus

C/Species: Caulobacter crescentus

C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001

C/Accession: F87700

R/Nieman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.

B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon

n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, D.C.; Fraser, C.M.

Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001

A/Title: Complete Genome Sequence of Caulobacter crescentus.

A/Reference number: AB7249; MUID:21173698; PMID:11259647

A/Accession: F87700

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-483 <STO>

A/Cross-references: GB:AE005673; NID:g13425394; PIDN:AAK25602.1; GSPDB:GN00148

C/Genetics:

A/Gene: CC3640

C/Superfamily: hypothetical protein b3052

Alignment Scores:

Pred. No.:	552	Length:	483
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.05%	Indels:	0
DB:	2	Gaps:	0

US-10-029-345A-108_COPY_532_2532 (1-2001) x F87700 (1-483)

OY 1829 ACTGGGGGAGCTGGCATG 1849

Db 239 ThrArgGlyAlaGlyMet 245

RESULT 48

B33501

myosin heavy chain 2, smooth muscle - rabbit (fragment)

C/Species: Oryctolagus cuniculus (domestic rabbit)

C/Date: 08-Dec-1989 #sequence_revision 08-Dec-1989 #text_change 13-Aug-1999

C/Accession: B33501; A31368

R/Nagel, R.; Kuro-o, M.; Babij, P.; Periasamy, M.

J. Biol. Chem. 264, 9734-9737, 1989

A/Title: Identification of two types of smooth muscle myosin heavy chain isoforms by cDN

A/Reference number: A33501; MUID:89255535; PMID:2722872

A/Accession: B33501

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-484 <NAG>

A/Cross-references: GB:J04833

R/Nagel, R.; Larson, D.M.; Periasamy, M.

Proc. Natl. Acad. Sci. U.S.A. 85, 1047-1051, 1988

A/Title: Characterization of a mammalian smooth muscle myosin heavy chain cDNA clone an

A/Reference number: A31368; MUID:88124972; PMID:3422477

A/Accession: A31368

A/Molecule type: mRNA

A/Residues: 1-484 <NAG>

A/Cross-references: GB:J03614; NID:g165517; PIDN:AAA31406.1; PID:g165518

C/Superfamily: myosin heavy chain; myosin motor domain homology

C/Keywords: muscle; smooth muscle

Pred. No.:	552	Length:	484
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.09%	Indels:	0
DB:	2	Gaps:	0

US-10-029-345A-108_COPY_532_2532 (1-2001) x B33501 (1-484)

OY 1403 CTGGCTCCCTTATCAGCA 1383

Db 263 LeuAlaSerLeuSerGly 269

RESULT 49

AF1107

phospho-beta-glucosidase homolog lmo0261 [imported] - Listeria monocytogenes (strain EGD

C/Species: Listeria monocytogenes

C/Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001

C/Accession: AF1107

R/Glaser, P.; Finguel, L.; Buchrieser, C.; Amend, A.; Baguerio, F.; Berche, P.; Bloeker

D.; Dominguez-Jernal, G.; Duchaud, E.; Durand, L.; Duesurget, O.; Entian, K.D.; Psitt, H.

Science 294, 849-852, 2001

A/Authors: Krefl, U.; Kunz, M.; Kunz, F.; Kurapat, G.; Mañeno, E.; Maitournam, A.; Ma

ok, C.; Schluter, T.; Simoes, N.; Tixeret, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,

A/Title: Comparative genomics of Listeria species.

A/Reference number: AB1077; MUID:21537279; PMID:11679669

A/Accession: AF1107

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-486 <GUA>

A:Cross-references: GB:NC_003210; PIDN:CAD00788.1; PID:gl6409626; GSPDB:GN00177
 A:Experimental source: strain EGD-e
 C:Genetics:
 A:Gene: lmo0261
 C:Superfamily: Agrobacterium beta-glucosidase

Alignment Scores:

Pred. No.:	552	Length:	486
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.09%	Indels:	0
DB:	2	Gaps:	0

US-10-029-345A-108_COPY_532_2532 (1-2001) x AP1107 (1-486)

QY 1791 ATGACTTGCTCGCGAAGT 1771

DB 434 lIeApIeuValSerAlaser 440

RESULT 50

AT1468

phospho-beta-glucosidase homolog lin0288 [imported] - Listeria innocua (strain Clp11262
 C:Species: Listeria innocua

C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #ext_change 14-Dec-2001

C:Accession: AT1468

R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,
 D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurgey, O.; Entian, K.D.; Fahl, H.

Science 294, 849-852, 2001

D.; Jones, L.M.; Karst, U.

A:Authors: Kreft, U.; Kuhn, M.; Kunst, F.; Kurapkai, G.; Madueno, E.; Maltournam, A.; Ma
 ok, C.; Schlueter, T.; Simoes, N.; Tlerez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehlend,

A:Title: Comparative genomics of Listeria species.

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AT1468

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-486 <GLA>

A:Cross-references: GB:AL592022; PIDN:CA095521.1; PID:gl6412717; GSPDB:GN00178

A:Experimental source: strain Clp11262

C:Genetics:

A:Gene: lin0288

C:Superfamily: Agrobacterium beta-glucosidase

Alignment Scores:

Pred. No.:	552	Length:	486
Score:	7.00	Matches:	7
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.09%	Indels:	0
DB:	2	Gaps:	0

US-10-029-345A-108_COPY_532_2532 (1-2001) x AT1468 (1-486)

QY 1791 ATGACTTGCTCGCGAAGT 1771

DB 434 lIeApIeuValSerAlaser 440

Search completed: June 21, 2004, 13:15:16
 Job time : 103.5 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 21, 2004, 12:46:21 ; Search time 22 Seconds
(without alignments)
9472.035 Million cell updates/sec

Title: US-10-029-345A-108_COPY_532_2532
Perfect score: 667
Sequence: 1 gttcgtcatgcccacatgagat.....tggaaatcatgaggtctcc 2001

Scoring table:
OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Word size: 1

Total number of hits satisfying chosen parameters: 66812

Minimum DB seq length: 473
Maximum DB seq length: 2000000000

Post-processing: Listing first 75 summaries

Command line parameters:
-MODE=frame+ n2p.model -DEV=xlp
-Q/cg2.1/USPTO.spool.p/US10029345/runat.21062004.124615.5680/app.query.fasta_1.2183
-DB=SwissProt.42 -QFMT=fastan -SUFFIX=oligo.rsp -MINMATCH=0.1 -LOOFC=0
-LOOEXT=0 -UNITS=bites -START=1 -END=-1 -MATRIX=oligo -TRANS=human0.cdi
-LIST=75 -DOCALLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=50 -MODE=LOCAL
-OUTFMT=ptc -NORM=exc -HEAPSIZE=500 -MINLEN=473 -MAXLEN=200000000
-USER=US10029345_@CGEN_1.1.22@runat.21062004.124615.5680 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	472	70.8	665	1 DUSG_HUMAN	Q9BY84 homo sapien
2	20	3.0	625	1 DUS8_HUMAN	Q13202 homo sapien
3	20	3.0	663	1 DUS8_MOUSE	O09112 mus musculu
4	9	1.3	619	1 VHP1_CABEL	Q10038 caenorhabdi
5	6	1.4	737	1 ACET_RABIT	P22968 oryctolagus
6	9	1.4	890	1 MUTS_RICCN	O92119 rickettsia
7	8	1.2	480	1 HEM3_EUGAR	P13446 euglena gira
8	8	1.2	509	1 AGP4_MOUSE	Q9WV16 mus musculu
9	8	1.2	526	1 VPS_BTVO1	P07389 bluetongue
10	8	1.2	555	1 VGLF_PIIHC	P12605 human para
11	8	1.2	605	1 GOX_ASPNG	P13006 aspergillus
12	8	1.2	640	1 ULJ5_HCVNA	P16766 human cytom
13	8	1.2	677	1 HUTN_CABEL	Q20502 caenorhabdi
14	8	1.2	835	1 VIRL_AGRU	P07167 agrobacteri
15	8	1.2	1716	1 RPA1_RAT	O54889 rattus norv
16	8	1.2	1717	1 RPA1_HUMAN	O95602 homo sapien
17	8	1.2	1717	1 RPA1_MOUSE	O35134 mus musculu
18	8	1.2	2512	1 POLN_SINDV	P03317 sindbis vir

ALIGNMENTS

19	8	1.2	2514	1 POLN_SINDV	P27283 sindbis vir
20	8	1.2	2842	1 APC_RAT	P70478 rattus norv
21	8	1.2	2843	1 APC_HUMAN	P25054 homo sapien
22	8	1.2	2845	1 APC_MOUSE	O61315 mus musculu
23	7	1.1	473	1 ABGA_CLOLO	O46130 clostridium
24	7	1.1	473	1 ATPB_PRES	O03080 pteridium e
25	7	1.1	474	1 ASCB_ECOLI	P24240 escherichia
26	7	1.0	475	1 PERI_MOUSE	P15331 mus musculu
27	7	1.1	475	1 SIM1_YEAST	P40472 saccharomyc
28	7	1.0	478	1 LXC2_PHOLE	P29236 photobacter
29	7	1.1	483	1 TR2_HUMAN	P13056 homo sapien
30	7	1.1	487	1 YMI8_STRAW	Q82K24 streptomyce
31	7	1.0	489	1 VGLY_MOPET	P19240 mopeia viru
32	7	1.1	491	1 YLID_YEAST	P40499 saccharomyc
33	7	1.1	494	1 FLIC_SALTY	P06179 salmonella
34	7	1.1	494	1 RNPL_HUMAN	Q9HAU8 homo sapien
35	7	1.1	496	1 HUTH_TREAC	Q9H116 thermoplasma
36	7	1.0	503	1 MATK_LIPO	O98370 liquidambar
37	7	1.0	503	1 MATK_LIPO	O84155 liquidambar
38	7	1.0	503	1 MATK_LIPO	O99053 liquidambar
39	7	1.0	507	1 MS14_ARATH	O22607 arabidopsis
40	7	1.0	510	1 PUR1_CHICK	P28173 gallus gall
41	7	1.0	511	1 HUTH_VIBCH	Q9KEG4 vibrio chol
42	7	1.1	517	1 COX1_MYXGL	O21079 myxine gluc
43	7	1.0	518	1 YH99_YEAST	P38758 saccharomyc
44	7	1.1	521	1 YAVB_SCHPO	O10177 schizosacch
45	7	1.0	524	1 HEMK_RICCN	Q92913 rickettsia
46	7	1.1	526	1 CAP_YEAST	P17555 saccharomyc
47	7	1.0	532	1 PPB_SCHPO	O60109 schizosacch
48	7	1.0	535	1 TCBI_AVEA	P40412 avena sativ
49	7	1.0	535	1 TCBI_AVEA	P54411 avena sativ
50	7	1.0	535	1 TCBI_AVEA	O04450 arabidopsis
51	7	1.1	540	1 CH60_THERB	O60024 thermotane
52	7	1.1	545	1 PYRG_VIBVU	O8DE33 vibrio vuln
53	7	1.1	548	1 CH60_OLEA	O8K30 oleispirita a
54	7	1.1	548	1 TCPO_HUMAN	P50390 homo sapien
55	7	1.1	548	1 TCPO_MOUSE	P42932 mus musculu
56	7	1.1	550	1 WR72_CABEL	O94130 caenorhabdi
57	7	1.1	551	1 CEH3_ECOLI	P00666 escherichia
58	7	1.0	559	1 GL65_CAPHI	P50426 capra hircu
59	7	1.0	564	1 NOX1_HUMAN	Q9Y588 homo sapien
60	7	1.0	566	1 SCCL_YEAST	O12158 saccharomyc
61	7	1.1	567	1 FIG1_HUMAN	O96929 homo sapien
62	7	1.0	567	1 IF37_SCHPO	O94236 schizosacch
63	7	1.0	574	1 CIRS_HUMAN	O43734 homo sapien
64	7	1.1	577	1 CHVG_RHIME	P72292 thizobium m
65	7	1.1	578	1 YDEM_CABEL	O19124 caenorhabdi
66	7	1.0	579	1 MALR_AEDAE	P13080 aedes aegy
67	7	1.0	583	1 SCB3_SCHPO	P1000 schizosacch
68	7	1.1	585	1 ACN3_CABEL	O9U745 caenorhabdi
69	7	1.0	585	1 ILV3_YEAST	P39522 saccharomyc
70	7	1.1	588	1 ATY1_MOUSE	O9CE66 mus musculu
71	7	1.0	591	1 UL49_EBV	P14347 Epstein-Bar
72	7	1.0	595	1 SIL1_HUMAN	O96921 homo sapien
73	7	1.1	596	1 CHVG_AGRIS	O07737 agrobacteri
74	7	1.0	597	1 SIL1_PANTR	O95180 pan troglod
75	7	1.0	597	1 Y745_ARATH	P59278 arabidopsis

RESULT 1

DUSG_HUMAN
ID DUSG_HUMAN STANDARD; PRT; 665 AA.
AC Q9BY84; Q9COG3; DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Dual specificity protein phosphatase 16 (EC 3.1.3.48) (EC 3.1.3.16)
DE (Mitogen-activated protein kinase phosphatase 7) (MKP-7)
DE phosphatase 7) (MKP-7).
GN DUSP16 OR MKP7 OR KIAA1700.
OS Homo sapiens (human).

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN R1
RP SEQUENCE FROM N.A.
RA MEDLINE=21486429; PubMed=11489891;
RX Masuda K., Shima H., Matanabe M., Kikuchi K.;
RT "MKP-7, a novel mitogen-activated protein kinase phosphatase,
RT functions as a shuttle protein.";
RL J. Biol. Chem. 276:39002-39011(2001).
RN R2
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=11214970;
RA Nagae T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIX.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:347-355(2000).
CC - FUNCTION: Involved in the inactivation of MAP kinases.
CC - CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H(2)O = protein
CC tyrosine + phosphate.
CC - CATALYTIC ACTIVITY: A phosphoprotein + H(2)O = a protein +
CC phosphate.
CC - SUBCELLULAR LOCATION: Cytoplasmic and nuclear.
CC - SIMILARITY: Belongs to the protein-tyrosine phosphatase family.
CC Non-receptor class dual specificity subfamily.
CC - SIMILARITY: Contains 1 rhodanese domain.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
-----
DR EMBL; AB052156; BAB40814.1; -.
DR EMBL; AB051487; BAB21791.1; ALT_INIT.
DR HSSP; Q16828; IMKP.
DR GeneW; HGNC:17909; DUSP16.
DR MIM; 607175; -.
DR GO; GO:0005737; C:cycloplasm; TAS.
DR GO; GO:0005634; C:nucleus; TAS.
DR GO; GO:0004721; F:protein phosphatase activity; TAS.
DR GO; GO:0016311; P:dephosphorylation; TAS.
DR GO; GO:0000188; P:inactivation of MAPK; TAS.
DR GO; GO:0043209; P:leptomycin B-sensitive MAPK phosphatase nuc. .; TAS.
DR GO; GO:0043204; P:MAPK nucleus export; TAS.
DR InterPro; IPR000340; DS_phosphatase.
DR InterPro; IPR008343; MAPK_phosph.
DR InterPro; IPR001763; Rhodanese-like.
DR InterPro; IPR000387; TYR_phosphatase.
DR Pfam; PF00782; DSPC; 1.
DR Pfam; PF00581; Rhodanese; 1.
DR PRINTS; PR01764; MARKPHPTASE.
DR SMART; SM00195; DSPC; 1.
DR SMART; SM00450; RHOD; 1.
DR PROSITE; PS50206; RHODANESE_3; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS50054; TYR_PHOSPHATASE_DUAL; 1.
KW Hydrolase; Nuclear protein.
FT DOMAIN 22 137 RHODANESE.
FT DOMAIN 227 289 PROTEIN-TYROSINE PHOSPHATASE.
FT ACT_SITE 244 244 PHOSPHOGLYCERATE INTERMEDIATE. (BY
SIMILARITY).
SQ SEQUENCE 665 AA; 73101 MW; 1BD853FF06460DF CRC64;
Alignment Scores:
Pred. No.: 0 Length: 665
Score: 472.00 Matches: 472
Percent Similarity: 100.00% Conservative: 0

```

Best Local Similarity:	100.00%	Mismatches:	0
Query Matchn:	70.76%	Indels:	0
DB:	1	Gaps:	0
US-10-029-345A-108_COPY_532_2532 (1-2001) x DUSG_HUMAN (1-665)			
QY	586	ACCTGTCCAAAGCTGATCTTATTCGCCGAGTCTCATTTCTCGGTGTCGCTGGATGAC	645
Db	194	ThrcysProlysbPpAapbPheilleProgluSerhIshPheulnarValProValAsnAsp	213
QY	646	AGCTTTGTGAGAAAATTTTGCCGTGGTGGACAAATCAGTAACTATTTCATTGAGAAAGCA	705
Db	214	SerPhecyGgluylsIleleuProtlPheAspLysSerValAspPheillegluylsAla	233
QY	706	AAAGCCTCCAAATGAGATGTGTTCTTAAGTCACATTTTACGTGGAGATCTCCGCTCCGACCC	765
Db	234	LysAlaserAsnDlyCysValleuValhIsCysleuAlaglylleSerhgsSerAlatThr	253
QY	766	ATGCGTATCGCTTACATCATGAGAGAGAGAGATGTCCTTACATGAGCTTACATTA	825
Db	254	IleAlaIleAlaIlyrlIleleclysArgMetAspSerleuAspLysValAspPheillegluylsAla	273
QY	826	GTCAAAAGAAAAGACCTTATATCTCCAACTTCAATTTCTCGGCCCACTCTGTGAC	885
Db	274	VallysgluylsArgPProThrIleSerProAsnPhAsnPhleuGllylIleuLeuAsp	293
QY	886	TATGAGAGAGAGATTAGAACCAAGATGGAGATCGAGGCCAAAGCAAACTCAAGCTG	945
Db	294	TygluIuylsYslIlelysAsnGlnItrgylAlaSerGlyProlySerIysleuylsleu	313
QY	946	CTGACCTGAGAGAGCCAAATGAACTGTCTCCGTGCTGTCTCAGAGGGGTGAGAGAAAAC	1001
Db	314	LeuHIsleuGlulysProAsnGluProValProAlaValSerGluGllylIleGlnlySer	333
QY	1006	GAAAGCCCTCAGATCCACCTGTGTGCCACTCTGCTCACTCAGAGGACAGACAAAGG	1067
Db	334	GlnThrProleuSerProCysAlaAspSerAlatThrSerGluAlaIleGlyGlnItrg	353
QY	1066	CCGTGCAATCCGCCACGCTGCCCCAGCGGCCAGCGTGCACCGCTCGCTGTAGAGAC	1122
Db	354	ProvalhIsProAlaserValProserValProserValGlnProSerleuLeuGlnAsp	373
QY	1126	AGCCGCTGTGACAGGCGCTCAGTGGGCTGCACCTGTCCGACAGACAGCTGGAAGACAC	1187
Db	374	SerProleuValGlnAlaleuSerGlyleuHIsleuSerAlaAspArgleuGluAspSer	393
QY	1186	AATAAGCTCAAGCGTCTCTCTCTCGGATATCAATAGTTTCATATTCGCGCAGCATG	1245
Db	394	AsnlybIeulysArgSerPheSerleuAspIlelySerSerValSerIlySerAlaserhIc	413
QY	1246	GCAGCATCTTACATGGCTTCTCTCATCAAGAAAGTCTTGGAACTATCAAACTTTC	1305
Db	414	AlaAlaserleuHIsGlyPheSerSerSerGluAspAlaIleuGluIlyTylybProSer	433
QY	1306	ACTATCTGGAGTGGGACCAACAAGTATGCCAGTCTCCCTGTTCAGAAATATCGGAG	1367
Db	434	ThrThrleuAspGlyThrAsnlybIeulysGlnPheSerProValGlnIleuSerIlu	453
QY	1366	CAGACTCCGAAACAGTCTGTATAGAGAGAGAGAGCAGCATCTCCCAAGAACTGCAAGAC	1425
Db	454	GlnThrProgluThrSerProAspIysIleGluAlaserIleProlyblybIeuglnThr	473
QY	1426	GCGAGGCTTTCAGACAGCCAGACAGAGGATTTGCATTGGTCAGAACACAGACAGCTGGC	1485
Db	474	AlaAspProSerAspSerGlnSerIlybArgleuHIsSerValIArgThrSerSerIoly	493
QY	1486	ACCGCCAGAGATGCCCTTTATCTCACTGATGAGAGTGGAGCGTGGAGAGCAATAC	1545
Db	494	ThrIlaGlnArgSerleuLeuSerProleuHIsArgSerGlySerValGluAspAsnItr	513
QY	1546	CACACAGACTCTTTTGCGCTTTTCCACAGCCAGACAGACCTCAGAACTGCTGGCC	1605
Db	514	HIsThrSerPheleuPheGlyleuSerThrSerGlnGlnIleuThrIlybSerAlagly	533


```
CC -1- SIMILARITY: Belongs to the protein-tyrosine phosphatase family.
CC Non-receptor class dual specificity subfamily.
CC -1- SIMILARITY: Contains 1 rhodanese domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X95518; CA64772.1; -.
CC HSSP; Q16828; IMKP.
CC MGD; MGI:106626; Dusp8.
CC InterPro; IPR000340; DS_phosphatase.
CC InterPro; IPR008343; MAPK_phosph.
CC InterPro; IPR001763; Rhodanese-like.
CC InterPro; IPR000387; TYR_phosphatase.
CC Pfam; PF00782; DSPC; 1.
CC PRINTS; PR01764; MAPKPHPTASE.
CC SMART; SM00195; DSPC; 1.
CC SMART; SM00450; RHOD; 1.
CC PROSITE; PS50206; RHODANES_3; 1.
CC PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
CC PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
CC PROSITE; PS50054; TYR_PHOSPHATASE_DUAL; 1.
CC K0 Hydroxylase; Nuclear protein.
CC FT DOMAIN 23 138 RHODANES.
CC FT DOMAIN 162 432 PROTEIN-TYROSINE PHOSPHATASE.
CC FT DOMAIN 452 459 POLY-ARG.
CC FT DOMAIN 555 558 POLY-SER.
CC FT DOMAIN 559 576 POLY-SER.
CC FT DOMAIN 577 600 POLY-GLY.
CC FT DOMAIN 311 552 PRO-RICH.
CC FT ACT_SITE 246 246 PHOSPHOCYSTEINE INTERMEDIATE (BY
SQ SEQUENCE 663 AA; 68847 MW; 416f429a12c1fa7c CRC64;
SIMILARITY).

Alignment Scores:
Pred. No.: 1,12e-11 Length: 663
Score: 20.00 Matches: 20
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.00% Indels: 0
DB: 1 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x DUS8_MOUSE (1-663)

OY 730 GTGCACCTGTTTACCTGGATCTCCGCTCCGACCATGCTATGCGCTACATCATGAAG 789
DB 244 VALHICysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIleMetLys 263

RESULT 4
VHP1_CAEEL
ID VHP1_CAEEL STANDARD; PRT; 619 AA.
AC 010038;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Protein-tyrosine phosphatase vhp-1 (EC 3.1.3.48).
GN VHP-1 OR F08B1.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Felodoridae; Caenorhabditis.
OX NCB1_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Chisoe S.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
CC tyrosine + phosphate.
```

```
CC -1- SIMILARITY: Belongs to the protein-tyrosine phosphatase family.
CC Non-receptor class dual specificity subfamily.
CC -1- SIMILARITY: Contains 1 rhodanese domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U23178; AAC6719.1; -.
CC PIR; T15969; T15969.
CC HSSP; Q16828; IMKP.
CC WormPep; F08B1.1; CE01899.
CC InterPro; IPR000340; DS_phosphatase.
CC InterPro; IPR000387; TYR_phosphatase.
CC Pfam; PF00782; DSPC; 1.
CC SMART; SM00195; DSPC; 1.
CC PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
CC PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
CC PROSITE; PS50054; TYR_PHOSPHATASE_DUAL; 1.
CC K0 Hydroxylase.
CC FT DOMAIN 200 269 PROTEIN-TYROSINE PHOSPHATASE.
CC FT DOMAIN 92 95 POLY-SER.
CC FT DOMAIN 351 354 POLY-SER.
CC FT DOMAIN 465 472 POLY-SER.
CC FT DOMAIN 483 488 POLY-SER.
CC FT ACT_SITE 224 224 PHOSPHOCYSTEINE INTERMEDIATE (BY
SQ SEQUENCE 619 AA; 66354 MW; 369E326F615D0529 CRC64;
SIMILARITY).

Alignment Scores:
Pred. No.: 2.61 Length: 619
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.35% Indels: 0
DB: 1 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x VHP1_CAEEL (1-619)

OY 733 CACTGTTTACCTGGATCTCCGCTCCGCTCC 759
DB 223 HICysLeuAlaGlyIleSerArgSer 231

RESULT 5
ACET_RABIT
ID ACET_RABIT STANDARD; PRT; 737 AA.
AC P22968;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Angiotensin-converting enzyme, testis-specific isoform precursor
DE Angiotensin-converting enzyme, testis-specific isoform precursor
DE (EC 3.4.15.1) (ACE-T) (Dipeptidyl carboxypeptidase I) (Kininase II).
GN ACE OR DCP1.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Butheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCB1_Taxid=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=New Zealand white; Tissue=Testis;
RC MEDLINE=89380303; PubMed=2550457;
RA Kumar R.S., Kuwari J., Roy S.N., Soffer R.L., Sen G.C.;
RT "Structure of testicular angiotensin-converting enzyme. A segmental
RT mosaic isozyme."
RL J. Biol. Chem. 264:16754-16758 (1989).
RN [2]
RP SEQUENCE OF 1-117 FROM N.A.
RC MEDLINE=91139683; PubMed=1847398;
RA Kumar R.S., Thekkumkara T.J., Sen G.C.;
RT "The mRNAs encoding the two angiotensin-converting isozymes are
```

transcribed from the same gene by a tissue-specific choice of
RT alternative transcription initiation sites.";
RL J. Biol. Chem. 266:3854-3862(1991).
CC
CC -1- FUNCTION: Converts angiotensin I to angiotensin II by release of
the terminal His-Leu, this results in an increase of the
vasoconstrictor activity of angiotensin.
CC
CC -1- CATALYTIC ACTIVITY: Release of a C-terminal dipeptide,
oligopeptide-|-Xaa-Xbb, when Xaa is not Pro, and Xbb is neither
Asp nor Glu. Converts angiotensin I to angiotensin II.
CC
CC -1- COFACTOR: Binds 1 zinc ion per subunit (By similarity).
CC
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC
CC -1- ALTERNATIVE PRODUCTS:
CC
CC Name=Testis-specific;
CC IsoId=P22368-1; Sequence=Displayed;
CC
CC Name=Somatic;
CC IsoId=P12822-1; Sequence=External;
CC
CC -1- TISSUE SPECIFICITY: Spermatoocytes, adult testis.
CC
CC -1- INDUCTION: Expression is thought to be subject to hormonal
regulation by androgens.
CC
CC -1- SIMILARITY: Belongs to peptidase family M2.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC
CC EMBL; J05041; AAA31153.1; -;
CC PIR; A34402; A34402. -;
CC
CC MEROPS; M02.004; -;
CC InterPro; IPR006025; Pept_M_Zn_BS.
CC InterPro; IPR001548; Peptidase_M2.
CC Pfam; PF01401; Peptidase_M2; 1.
CC PRINTS; PR00791; PEPTIDPTSEA.
CC PRODOM; PD004184; Peptidase_M2; 1.
CC DR PROSITE; PS00142; ZINC_PROTEASE; 1.
CC Hydrolase; Metalloprotease; Carboxypeptidase; Zinc; Glycoprotein;
CC Transmembrane; Testis; Signal; Alternative splicing.
CC
CC FT CHAIN 1 32 ANGIOTENSIN-CONVERTING ENZYME, TESTIS-
FT SIGNAL 33 737 SPECIFIC ISOFORM.
FT DOMAIN 33 690 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 691 707 POTENTIAL.
FT DOMAIN 708 737 CYTOPLASMIC (POTENTIAL).
FT METAL 419 419 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT SITE 420 420 BY SIMILARITY.
FT METAL 423 423 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 447 447 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 188 194 BY SIMILARITY.
FT DISULFID 388 406 BY SIMILARITY.
FT DISULFID 574 586 BY SIMILARITY.
FT CARBOHYD 108 108 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 145 145 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 373 373 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 622 622 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 737 AA; 83923 MW; FC43CCTG6553DCA CRC64;

Alignment Scores:
Pred. No.: 2.56 Length: 737
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.40% Indels: 0
DB: 1 Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x ACET_RABIT (1-737)
QY 1334 CATAGCTTGTGGTCCATCCAGTA 1308

Db 22 HisSerLeuValProSerArgVal 30
|||||
RESULT 6
ID MUTS_RICCN STANDARD; PRT; 890 AA.
AC 092119;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE DNA mismatch repair protein muts.
GN MUTS OR RCO401.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
OC Rickettsiaceae; Rickettsiaseae; Rickettsia.
OX NCBI_TaxId=781;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Mallish 7;
RC MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-F., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
CC
CC -1- FUNCTION: This protein is involved in the repair of mismatches in
CC DNA. It is possible that it carries out the mismatch recognition
CC step. This protein has a weak ATPase activity (By similarity).
CC
CC -1- SIMILARITY: Belongs to the DNA mismatch repair muts family.

CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC
CC EMBL; AE008604; ALN02939.1; -;
CC PIR; A97750; A97750.
CC HAMAP; MF_00096; -; 1.
CC InterPro; IPR005748; MutS1.
CC InterPro; IPR000432; MutS_C.
CC InterPro; IPR007860; MutS_I.
CC InterPro; IPR007866; MutS_III.
CC InterPro; IPR007896; MutS_IV.
CC InterPro; IPR007861; MutS_N.
CC Pfam; PF01624; MutS_I; 1.
CC Pfam; PF05188; MutS_II; 1.
CC Pfam; PF05192; MutS_III; 1.
CC Pfam; PF05190; MutS_IV; 1.
CC Pfam; PF00488; MutS_V; 1.
CC PRODOM; PD001263; MutS_C; 1.
CC DR SMART; SM00534; MUTSac; 1.
CC DR SMART; SM00533; MUTSac; 1.
CC TIGRfam; TIGR01070; mutS1; 1.
CC DR PROSITE; PS00486; DNA_MISMATCH_REPAIR_2; 1.
CC DR DNA repair; ATP-binding; DNA-binding; Complete proteome.
FT NP BIND 645 652 ATP (POTENTIAL).
SQ SEQUENCE 890 AA; 100376 MW; 5016F26921D16792 CRC64;

Alignment Scores:
Pred. No.: 2.5 Length: 890
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.40% Indels: 0
DB: 1 Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x MUTS_RICCN (1-890)
QY 1406 ATGCTGCTTCTCTTATCAGACTG 1380
|||||

DB 572 MetLeuAlaSerSerLeuSerGlyLeu 580

RESULT 7

ID	HEM3_EUGGR	STANDARD	PRT	480 AA.
AC	P13446			
DT	01-JAN-1990	(Rel. 13, Created)		
DT	01-JAN-1990	(Rel. 13, Last sequence update)		
DT	15-MAR-2004	(Rel. 43, Last annotation update)		
DE	Porphobilinogen deaminase, chloroplast precursor (EC 2.5.1.61) (PBG)			
DE	(Hydroxymethylbilane synthase) (HMB) (Pre-urophorphyrinogen synthase).			
OS	Euglena gracilis.			
OC	Eukaryota; Euglenozoa; Euglenida; Euglenales; Euglena.			
OX	NCBI_TaxID=3039;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RC	STRAIN=Z / UTEX 753.			
RX	MEDLINE=90005485; PubMed=2477247;			
RA	Sharif A.U., Smith A.G., Abell C.;			
RT	"Isolation and characterization of a cDNA clone for a chlorophyll			
RT	synthesis enzyme from Euglena gracilis. The chloroplast enzyme			
RT	hydroxymethylbilane synthase (porphobilinogen deaminase) is			
RT	synthesized with a very long transit peptide in Euglena."			
RL	Eur. J. Biochem. 194:353-359(1989).			
CC	-1- FUNCTION: tetrapolymerization of the monopyrrole PBG into the			
CC	hydroxymethylbilane precursorporphyrinogen in several discrete steps.			
CC	-1- CATALYTIC ACTIVITY: 4 porphobilinogen + H(2)O =			
CC	hydroxymethylbilane + 4 NH(3).			
CC	-1- COFACTOR: Covalently binds a dipyrromethane cofactor to which the			
CC	porphobilinogen subunits are added.			
CC	-1- PATHWAY: Porphyrin biosynthesis by the C5 pathway; fourth step.			
CC	-1- INVOLVED in chlorophyll biosynthesis.			
CC	-1- SIMILARITY: Belongs to the HMBs family.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sb-sib.ch).			
CC	-----			
DR	EMBL: X15743; CAA33759.1; -.			
DR	PIR: S06109; IBBG.			
DR	HSSP: P06983; IPDA.			
DR	InterPro: IPR000860; Porphobil deam.			
DR	Pfam: PF01379; Porphobil deamC. 1.			
DR	Pfam: PF03900; Porphobil deamC. 1.			
DR	PRINTS: PR00151; PORPHBDMASE.			
DR	ProDom: PD002745; Porphobil deam; 1.			
DR	TIGRPFAM: TIGR00212; hmcC_2.			
DR	PROSITE: PS00533; PORPHOBILINOGEN DEAM; 1.			
KW	Porphyrin biosynthesis; Chlorophyll biosynthesis; Transferase;			
KW	Chloroplast; Transit peptide.			
FT	TRANSIT 1 139 CHLOROPLAST.			
FT	CHAIN 140 480 PORPHOBILINOGEN DEAMINASE.			
FT	BINDING 395 395 PYRROMETHANE COFACTOR (BY SIMILARITY).			
FT	SEQUENCE 480 AA; 51743 MW; 269CE6CC195C0F3A CRC64;			
Pred. No.:	29	Length:	480	
Score:	8.00	Matches:	8	
Percent Similarity:	100.00%	Conservative:	0	
Best Local Similarity:	100.00%	Mismatches:	0	
Query Match:	1.20%	Indels:	0	
DB:	1	Gaps:	0	

RESULT 8

ID	AGP4_MOUSE	STANDARD	PRT	509 AA.
AC	Q9VWH6			
DT	28-FEB-2003	(Rel. 41, Created)		
DT	28-FEB-2003	(Rel. 41, Last sequence update)		
DT	28-FEB-2003	(Rel. 41, Last annotation update)		
DE	Angiopoietin-4 precursor (ANG-4) (ANG-3).			
GN	ANGPT4 OR AGPT4 OR ANG3.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Myoblasts, and uterus;			
RX	MEDLINE=99162530; PubMed=10051567;			
RA	Valenzuela D.M., Griffiths J.A., Rojas J., Aldrich T.H., Jones P.F.,			
RA	Zhou H., McClain J., Copeland N.G., Gilbert D.J., Jenkins N.A.,			
RA	Huang T., Papadopoulos N., Malsomperre P.C., Davis S.,			
RT	Yancopoulos G.D.;			
RT	"Angiopoietins 3 and 4: diverging gene counterparts in mice and			
RT	humans."			
RL	Proc. Natl. Acad. Sci. U.S.A. 96:1904-1909(1999).			
CC	-1- FUNCTION: BINDS TO TYROSINE-PROTEIN KINASE RECEPTOR TIE2 AND MAY			
CC	ACT AS AN ANTAGONIST.			
CC	-1- SUBCELLULAR LOCATION: Secreted (Probable).			
CC	-1- TISSUE SPECIFICITY: Widely expressed.			
CC	-1- SIMILARITY: Contains 1 fibrinogen C-terminal domain.			
CC	-1- CAUTION: Was originally (Ref.1) called angiopoietin-3.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sb-sib.ch).			
CC	-----			
DR	EMBL: AF113707; AAD21586.1; -.			
DR	HSSP: P02671; IF2D.			
DR	MGI: MGI:1336887; Agp4.			
DR	InterPro: IPR002181; Fibrinogen_C.			
DR	Pfam: PF00147; fibrinogen_C; 1.			
DR	SMART: SM00186; FBG; 1			
DR	PROSITE: PS00514; FIBRIN_AG_C DOMAIN; 1.			
KW	Coiled coil; Signal.			
FT	SIGNAL 1 21 POTENTIAL.			
FT	CHAIN 22 509 ANGIOPOIETIN-4.			
FT	DOMAIN 181 269 COILED COIL (POTENTIAL).			
FT	DOMAIN 294 480 FIBRINOGEN-LIKE.			
FT	DISULFD 297 326 BY SIMILARITY.			
FT	DISULFD 450 463 BY SIMILARITY.			
FT	DISULFD 450 463 BY SIMILARITY.			
SO	SEQUENCE 509 AA; 57805 MW; 9B5A74A20A664F6 CRC64;			
Pred. No.:	28.8	Length:	509	
Score:	8.00	Matches:	8	
Percent Similarity:	100.00%	Conservative:	0	
Best Local Similarity:	100.00%	Mismatches:	0	
Query Match:	1.20%	Indels:	0	
DB:	1	Gaps:	0	
ID	VPS_BT10	STANDARD	PRT	526 AA.
AC	P07389			

DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Outer capsid protein VP5.
GN 56 OR M5.
OS Bluesongue virus (serotype 10 / isolate USA).
OC Viruses; dsRNA viruses; Reoviridae; Orbivirus.
NCBI_TaxID=10900;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=6198636; PubMed=3009695;
RA Purdy M.A., Ritter G.D., Roy P.;
RT "Nucleotide sequence of cDNA clones encoding the outer capsid protein, VP5, of bluesongue virus serotype 10.";
RT J. Gen. Virol. 67:957-962(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90345726; PubMed=2166648;
RA Roy P., Marshall J.J.A., French T.J.;
RT "Structure of the bluesongue virus genome and its encoded proteins.";
RT Curr. Top. Microbiol. Immunol. 162:43-87(1990).
CC -1- FUNCTION: THE VP5 PROTEIN IS ONE OF THE TWO PROTEINS (WITH VP2) WHICH CONSTITUTE THE VIRUS PARTICLE OUTER CAPSID.
CC -1- SIMILARITY: Belongs to the reoviruses VP5 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.1sb-sib.ch/announce/> or send an email to license@1sb-sib.ch).
CC -----
CC EMBL; D12532; BAA02095.1; -.
DR PIR; A25419; PSXR10.
DR InterPro; IPR000145; Orbi_VP5.
DR Pfam; PF00901; Orbi_VP5; 1.
KM Coat protein.
SQ SEQUENCE 526 AA; 59162 MW; 29AE556BDB55522C CRC64;
SO
Alignment Scores:
Pred. No.: 28.7 Length: 526
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0
DB: 1 Gaps: 0
US-10-029-345A-108_COPY_532_2532 (1-2001) x VP5_BT10 (1-526)
QY 98 CTATCAATTGCGACACTTTTCC 75
DB 435 LeuSerIleSerSerThrPheSer 442
RESULT 10
VGLF_P1IHC STANDARD; PRT; 555 AA.
AC P12605;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fusion glycoprotein precursor [Contains: Fusion glycoprotein F2; DE Fusion glycoprotein F1].
GN F.
OS Human parainfluenza 1 virus (strain C39).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Respirovirus.
NCBI_TaxID=11210;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89045674; PubMed=2847427;
RA Merzon J.R., Hull R.A., Bates M.K., Kaseel J.A.;
RT "Molecular cloning and sequence determination of the fusion protein

RT Gene of human parainfluenza virus type 1.";
RL Virology 167:97-105(1988).
CC -1- FUNCTION: This protein directs fusion of viral and cellular membranes.
CC -1- SUBUNIT: THE MATURE FORM IS A DIMER OF POLYPEPTIDES F-1 AND F-2 LINKED BY A DISULFIDE BOND.
CC -1- SIMILARITY: Belongs to the paramyxoviruses fusion glycoprotein family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.1sb-sib.ch/announce/> or send an email to license@1sb-sib.ch).
CC -----
CC EMBL; M22347; AAA46800.1; -.
DR PIR; A13287; VGN211.
DR InterPro; IPR000776; Fusion_gly.
DR Pfam; PF00523; Fusion_gly; 1.
KM Glycoprotein; Fusion protein; Transmembrane; Envelope protein; Signal.
FT FT SIGNAL 1 21
FT CHAIN 22 555
FT CHAIN 22 112
FT DISULFID 66 195
FT TRANSMEM 113 138
FT TRANSMEM 498 519
FT CARBOHYD 100 100
FT CARBOHYD 241 241
FT CARBOHYD 529 529
FT CARBOHYD 552 552
SQ SEQUENCE 555 AA; 60785 MW; 6F0DF6E2C969B9F0 CRC64;
SO
Alignment Scores:
Pred. No.: 28.5 Length: 555
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0
DB: 1 Gaps: 0
US-10-029-345A-108_COPY_532_2532 (1-2001) x VGLF_P1IHC (1-555)
QY 1193 AGCTTATGCTGTCTTCAGCCTG 1170
DB 13 SerLeuLeuSerSerSerLeu 20
RESULT 11
GOX_ASPNG STANDARD; PRT; 605 AA.
ID GOX_ASPNG
AC P13006;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Glucose oxidase precursor (EC 1.1.3.4) (Glucose oxyhydrase) (GOD) DE (Beta-D-glucose:oxygen 1-oxido-reductase).
GN GOX.
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
NCBI_TaxID=5061;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=NRRL 3;
RX MEDLINE=90005974; PubMed=2792372;
RA Kriechbaum M., Hellmann H.J., Wientjes F.J., Hahn M., Jany K.D., Gassen H.G., Sharif F., Alaeddingu G.;
RT "Cloning and DNA sequence analysis of the glucose oxidase gene from Aspergillus niger NRRL-3.";
RT FEBS Lett. 255:63-66(1989).
RN [2]

RP SEQUENCE FROM N.A.
RX MEDLINE=90154060; PubMed=2406261;
RA Frederick K.R., Tung J., Emerick R.S., Mastarz F.R., Chamberlain S.H.,
RA Vasavada A., Rosenberg S., Chakraborty S., Schopler L.M., Massey V.,
RT "Glucose oxidase from *Aspergillus niger*. Cloning, gene sequence,
RT secretion from *Saccharomyces cerevisiae* and kinetic analysis of a
RT yeast-derived enzyme.";
RL J. Biol. Chem. 265:3793-3802(1990).
RN [3]
RP SEQUENCE OF 1-32 FROM N.A.
RX MEDLINE=91168301; PubMed=2076553;
RA Whittington H., Kerry-Williams S., Bidgood K., Dodsworth N.,
RA Pedery J., Dobson M., Hinchliffe E., Ballance D.J.,
RT "Expression of the *Aspergillus niger* glucose oxidase gene in *A.*
RT *niger*, *A. nidulans* and *Saccharomyces cerevisiae*.";
RL Curr. Genet. 18:531-536(1990).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE=9133782; PubMed=8421298;
RA Hecht H.-J., Kalisz H.M., Hendle J., Schmid R.D., Schomburg D.,
RT "Crystal structure of glucose oxidase from *Aspergillus niger* refined
RT at 2.3-A resolution.";
RL J. Mol. Biol. 229:153-172(1993).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE=99234348; PubMed=10216293;
RA Wohlfahrt G., Wilt S., Hendle J., Schomburg D., Kalisz H.M.,
RA Hecht H.-J.,
RT "1.8 and 1.9-A resolution structures of the Penicillium amagaaklense
RT and *Aspergillus niger* glucose oxidases as a basis for modelling
RT substrate complexes.";
RL Acta Crystallogr. D 55:969-977(1999).
CC -1- CATALYTIC ACTIVITY: Beta-D-glucose + O(2) = D-glucono-1,5-lactone
CC + H(2)O(2).
CC -1- COFACTOR: FAD.
CC -1- SUBUNIT: Homodimer.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MISCELLANEOUS: THIS ENZYME IS WIDELY APPLIED FOR THE DETERMINATION
CC OF GLUCOSE IN BODY FLUIDS AND IN REMOVING RESIDUAL GLUCOSE OR
CC OXYGEN FROM FOODS AND BEVERAGES. FURTHERMORE, GLUCOSE
CC OXIDASE-PRODUCING MOUTHS SUCH AS *ASPERGILLUS* AND *PENICILLIUM*
CC SPECIES ARE USED FOR THE BIOLOGICAL PRODUCTION OF GLUCONIC ACID.
CC -1- SIMILARITY: Belongs to the GMC oxidoreductase family.
CC -1- DATABASE: NAME=Worthington enzyme manual;
CC WWW="http://www.worthington-biochem.com/cop/".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X16061; CAA34197.1; -;
CC EMBL: J05242; AAA32695.1; -;
CC EMBL: X56443; CAA39826.1; -;
CC PIR: A35459; A35459.
CC PDB: 1GAL; 31-OCT-93.
CC PDB: 1CF3; 26-MAR-99.
CC InterPro: IPR000172; GMC_oxred.
CC InterPro: IPR007867; GMC_oxred_C.
CC Pfam: PF05199; GMC_oxred_C; 1.
CC Pfam: PF00732; GMC_oxred_N; 1.
CC PROSITE: PS00623; GMC_OXRED_1; 1.
CC PROSITE: PS00624; GMC_OXRED_2; 1.
CC Oxidoreductase; Flavo-protein; FAD; Glycoprotein; Signal; 3D-structure.
FT SIGNAL 1 22
FT CHAIN 23 605
FT NP_BIND 43 72
FT ACT_SITE 543 543
FT DISULFID 186 228
FT CARBOHYD 111 111
FT N-LINKED (GLCNAC. . .).
FT

FT HELIX 26 29
FT STRAND 31 31
FT HELIX 34 36
FT TURN 37 39
FT STRAND 41 47
FT HELIX 51 60
FT TURN 61 62
FT TURN 64 65
FT STRAND 68 72
FT TURN 78 79
FT HELIX 82 85
FT STRAND 86 86
FT HELIX 87 89
FT TURN 90 95
FT TURN 97 98
FT STRAND 99 99
FT STRAND 102 103
FT TURN 108 110
FT STRAND 115 116
FT STRAND 119 119
FT TURN 122 123
FT HELIX 124 127
FT TURN 128 128
FT STRAND 132 132
FT HELIX 138 146
FT TURN 147 148
FT TURN 151 152
FT HELIX 155 165
FT STRAND 166 168
FT HELIX 173 178
FT TURN 179 179
FT HELIX 184 186
FT STRAND 193 196
FT STRAND 204 204
FT TURN 206 206
FT HELIX 207 216
FT TURN 217 219
FT STRAND 222 222
FT TURN 226 227
FT STRAND 233 235
FT STRAND 239 239
FT STRAND 241 241
FT TURN 243 244
FT STRAND 246 247
FT HELIX 250 254
FT TURN 255 255
FT HELIX 256 258
FT TURN 259 260
FT TURN 262 263
FT STRAND 264 268
FT STRAND 271 278
FT STRAND 285 294
FT TURN 295 296
FT STRAND 297 309
FT TURN 313 315
FT HELIX 316 322
FT TURN 323 324
FT STRAND 325 325
FT HELIX 328 331
FT STRAND 332 334
FT STRAND 335 335
FT TURN 340 340
FT STRAND 344 345
FT STRAND 348 349
FT STRAND 352 360
FT HELIX 362 364
FT STRAND 367 367
FT STRAND 369 375
FT STRAND 376 380
FT HELIX 381 383
FT HELIX 384 393
FT HELIX 395 404
FT TURN 405 406

FT HELIX 411 426
FT TURN 427 428
FT STRAND 431 438
CC 440 441
FT TURN 442 449
FT STRAND 446 460
FT HELIX 465 467
FT STRAND 470 473
FT TURN 476 477
FT HELIX 480 497
FT TURN 498 498
FT HELIX 500 504
FT TURN 505 505
FT STRAND 506 511
FT HELIX 513 515
FT TURN 518 519
FT HELIX 522 528
FT HELIX 529 531
FT STRAND 534 535
FT STRAND 540 540
FT TURN 542 543
FT STRAND 545 545
FT HELIX 548 550
FT TURN 551 551
FT STRAND 554 554
FT TURN 555 557
FT STRAND 559 560
FT TURN 561 562
FT STRAND 564 564
FT STRAND 566 568
FT TURN 571 572
FT HELIX 583 603

Alignment Scores:

Pred. No.: 28.2 Length: 605
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 1 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x GOX_AS PNG (1-605)

QY 386 AGTCTCTGCTGTTCCCTGGCC 409

DB 7 SerSerLeuValValSerLeuAla 14

RESULT 12

UL35_HCMVA STANDARD; PRT; 640 AA.

AC P16766; DT 01-AUG-1990 (Rel. 15, Created)

DT 01-AUG-1990 (Rel. 15, Last sequence update)

DT 01-FEB-1991 (Rel. 17, Last annotation update)

DE Hypothetical protein UL35.

GN UL35.

OS Human cytomegalovirus (strain AD169).

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Betaherpesvirinae; Cytomegalovirus.

OX NCBI_TaxID=10360; [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=90269039; PubMed=2161319; Beck S., Brown C.M., Cerny R.,

RA Chee W.S., Bankier A.T., Horensell T., Hutchinson C.A. III, Kozarides T., Martignetti J.A.,

RA Peddie E., Satchwell S.C., Tomlinson P., Weston K.M., Barrett B.G.,

RT "Analysis of the protein-coding content of the sequence of human

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X17403; CAA35394.1; -
DR PIR; S09798; QOBEU2.
DR InterPro; IPR006731; Herpes_pp85.
DR Pfam; PF04637; Herpes_pp85; 1.
KW Hypothetical protein.
FT CARBOHYD 208 208 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 640 AA; 72529 MW; 8FCCFA3C9F631C1C CRC64;

Alignment Scores:

Pred. No.: 28.1 Length: 640
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0
DB: 1 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x UL35_HCMVA (1-640)

QY 1190 TTATGCTGCTTCCAGCCGCT 1167

DB 547 LeuLeuLeuSerSerLeuSer 554

RESULT 13

HUTH CAEEL STANDARD; PRT; 677 AA.

AC 020502; DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DE Probable histidine ammonia-lyase (EC 4.3.1.3) (Histidase).

GN F47810.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI_TaxID=6239; [1]

RP SEQUENCE FROM N.A.

RA Harris B.; DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DE Probable histidine ammonia-lyase (EC 4.3.1.3) (Histidase).

GN F47810.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI_TaxID=6239; [1]

RP SEQUENCE FROM N.A.

RA Harris B.; DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DE Probable histidine ammonia-lyase (EC 4.3.1.3) (Histidase).

GN F47810.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI_TaxID=6239; [1]

RP SEQUENCE FROM N.A.

RA Harris B.; DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DE Probable histidine ammonia-lyase (EC 4.3.1.3) (Histidase).

GN F47810.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI_TaxID=6239; [1]

RP SEQUENCE FROM N.A.

RA Harris B.; DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DE Probable histidine ammonia-lyase (EC 4.3.1.3) (Histidase).

GN F47810.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI_TaxID=6239; [1]

RP SEQUENCE FROM N.A.

RA Harris B.; DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DE Probable histidine ammonia-lyase (EC 4.3.1.3) (Histidase).

GN F47810.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI_TaxID=6239; [1]

RP SEQUENCE FROM N.A.

RA Harris B.; DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DE Probable histidine ammonia-lyase (EC 4.3.1.3) (Histidase).

GN F47810.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI_TaxID=6239; [1]

RP SEQUENCE FROM N.A.

RA Harris B.; DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DE Probable histidine ammonia-lyase (EC 4.3.1.3) (Histidase).

GN F47810.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI_TaxID=6239; [1]

RP SEQUENCE FROM N.A.

RA Harris B.; DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DE Probable histidine ammonia-lyase (EC 4.3.1.3) (Histidase).

GN F47810.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI_TaxID=6239; [1]

RP SEQUENCE FROM N.A.

RA Harris B.; DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DE Probable histidine ammonia-lyase (EC 4.3.1.3) (Histidase).

GN F47810.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI_TaxID=6239; [1]

RP SEQUENCE FROM N.A.

RA Harris B.; DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DE Probable histidine ammonia-lyase (EC 4.3.1.3) (Histidase).

GN F47810.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI_TaxID=6239; [1]

RP SEQUENCE FROM N.A.

RA Harris B.; DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DE Probable histidine ammonia-lyase (EC 4.3.1.3) (Histidase).

GN F47810.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI_TaxID=6239; [1]

RP SEQUENCE FROM N.A.

RA Harris B.; DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DE Probable histidine ammonia-lyase (EC 4.3.1.3) (Histidase).

GN F47810.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI_TaxID=6239; [1]

RP SEQUENCE FROM N.A.

RA Harris B.; DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DE Probable histidine ammonia-lyase (EC 4.3.1.3) (Histidase).

GN F47810.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI_TaxID=6239; [1]

RP SEQUENCE FROM N.A.

RA Harris B.; DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DE Probable histidine ammonia-lyase (EC 4.3.1.3) (Histidase).

GN F47810.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI_TaxID=6239; [1]

RP SEQUENCE FROM N.A.

RA Harris B.; DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DE Probable histidine ammonia-lyase (EC 4.3.1.3) (Histidase).

GN F47810.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI_TaxID=6239; [1]

RP SEQUENCE FROM N.A.

RA Harris B.; DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DE Probable histidine ammonia-lyase (EC 4.3.1.3) (Histidase).

GN F47810.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI_TaxID=6239; [1]

RP SEQUENCE FROM N.A.

RA Harris B.; DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DE Probable histidine ammonia-lyase (EC 4.3.1.3) (Histidase).

GN F47810.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI_TaxID=6239; [1]

RP SEQUENCE FROM N.A.

RA Harris B.; DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DE Probable histidine ammonia-lyase (EC 4.3.1.3) (Histidase).

GN F47810.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI_TaxID=6239; [1]

RP SEQUENCE FROM N.A.

RA Harris B.; DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DE Probable histidine ammonia-lyase (EC 4.3.1.3) (Histidase).

GN F47810.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI_TaxID=6239; [1]

RP SEQUENCE FROM N.A.

RA Harris B.; DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DE Probable histidine ammonia-lyase (EC 4.3.1.3) (Histidase).

GN F47810.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI_TaxID=6239; [1]

RP SEQUENCE FROM N.A.

RA Harris B.; DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DE Probable histidine ammonia-lyase (EC 4.3.1.3) (Histidase).

GN F47810.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI_TaxID=6239; [1]

RP SEQUENCE FROM N.A.

RA Harris B.; DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DE Probable histidine ammonia-lyase (EC 4.3.1.3) (Histidase).

GN F47810.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI_TaxID=6239; [1]

RP SEQUENCE FROM N.A.

RA Harris B.; DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DE Probable histidine ammonia-lyase (EC 4.3.1.3) (Histidase).

GN F47810.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI_TaxID=6239; [1]

RP SEQUENCE FROM N.A.

RA Harris B.; DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DE Probable histidine ammonia-lyase (EC 4.3.1.3) (Histidase).

GN F47810.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI_TaxID=6239; [1]

RP SEQUENCE FROM N.A.

RA Harris B.; DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DE Probable histidine ammonia-lyase (EC 4.3.1.3) (Histidase).

GN F47810.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI_TaxID=6239; [1]

FT MOD_RES 270 270 (By similarity).
FT SEQ 677 AA; 74634 MW; B64CCD5B097AAC4F CRC64;
SQ SEQUENCE Alignment Scores:
Pred. No.: 27.9 Length: 677
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 1 Gaps: 0
US-10-029-345A-108_COPY_532_2532 (1-2001) x HUTH_CABEL (1-677)
QY 137 AAGCATTATATCAATGCTCCA 160
DB 584 LysPleuileSerThrAlaPro 591
RESULT 14
VIRL_AGRU STANDARD; PRT; 835 AA.
AC P07167;
DT 01-APR-1998 (Rel. 07, Created)
DT 01-APR-1998 (Rel. 07, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Limited host range virA protein (EC 2.7.3.-) (LHR virA).
GN VIRA.
OS Agrobacterium tumefaciens.
OC Plasmid pTiA6.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=358;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=A348, and A856;
RX MEDLINE=87246518; PubMed=3595559.
RA Leroux B., Yanozsky M.F., Winans S.C., Ward J.E., Ziegler S.F.,
RA Neeter E.W.;
RT "Characterization of the virA locus of Agrobacterium tumefaciens: a
RT transcriptional regulator and host range determinant.";
RL EMO J. 6:849-856(1987).
CC -1- SIMILARITY: Contains 1 histidine kinase domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-stb.ch/announce/>
CC or send an email to license@isb-stb.ch).
CC -----
DR EMBL; X05241; CAA28868.1; -
DR PIR; A27211; A27211.
DR InterPro; IPR003594; ATPbind_Artpase.
DR InterPro; IPR004358; Bact_sens_pr_C.
DR InterPro; IPR003661; His_kinA_N.
DR InterPro; IPR005467; His_kinase.
DR InterPro; IPR001789; Response_reg.
DR Pfam; PF02518; HATPase_c_1.
DR Pfam; PF00512; HisKA_1.
DR PRINTS; PR00344; BCTRLSENSOR.
DR SMART; SM00387; HATPase_c_1.
DR SMART; SM00388; HisKA_1.
DR SMART; SM00448; REC_1.
DR PROSITE; PS50109; HIS_KIN_1.
KW Sensory transduction; Transferase; Kinase; Phosphorylation;
KW PfamId; Crown_gall_tumor; Transmembrane.
FT TRANSMEM 19 37
FT DOMAIN 476 698 HISTIDINE KINASE.
FT MOD_RES 479 479 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
SQ SEQUENCE 835 AA; 92443 MW; 93ADBA1E2B5F801 CRC64;

Alignment Scores:
Pred. No.: 27.2 Length: 835
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0
DB: 1 Gaps: 0
US-10-029-345A-108_COPY_532_2532 (1-2001) x VIRL_AGRU (1-835)
QY 545 ATGAGCTCTGTGTGAGCATCT 522
DB 148 IleSerSerLeuLeuArgThrSer 155
RESULT 15
RPA1_RAT
ID_RPA1_RAT STANDARD; PRT; 1716 AA.
AC 0548R9;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA-directed RNA polymerase I largest subunit (EC 2.7.7.6) (RPA
DE polymerase I 194 kDa subunit) (RPA194).
GN POLR1A OR RPO1-4 OR RPA1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98086309; PubMed=9422795;
RX Hannan R.D., Hempel W.M., Cavanaugh A., Arino T., Dimitrov S.I.,
RA Moses T., Kochdum L.;
RT "Affinity purification of mammalian RNA polymerase I. Identification
RT of an associated kinase.";
RL J. Biol. Chem. 273:1257-1267(1998).
CC -1- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription
CC of DNA into RNA using the four ribonucleoside triphosphates as
CC substrates. RNA polymerase I is essentially used to transcribe
CC ribosomal DNA units.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC -1- SUBUNIT: Each class of RNA polymerase is assembled from 9 to 14
CC different polypeptides. This subunit is the largest component of
CC RNA polymerase I.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- MISCELLANEOUS: Three distinct zinc-containing RNA polymerases are
CC found in eukaryotic nuclei: polymerase I for the ribosomal RNA
CC precursor, polymerase II for the mRNA precursor, and polymerase
CC III for 5S and tRNA genes.
CC -1- SIMILARITY: Belongs to the RNA polymerase beta' chain family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-stb.ch/announce/>
CC or send an email to license@isb-stb.ch).
CC -----
DR EMBL; AF025425; AAB94601.1; -
DR PIR; T14103; T14103.
DR InterPro; IPR000722; RNA_pol_A.
DR InterPro; IPR007080; RNA_pol_Rpb1_1.
DR InterPro; IPR007066; RNA_pol_Rpb1_3.
DR InterPro; IPR007083; RNA_pol_Rpb1_4.
DR InterPro; IPR007081; RNA_pol_Rpb1_5.
DR InterPro; IPR006592; RNA_pol_N.
DR Pfam; PF04997; RNA_pol_Rpb1_1_1.
DR Pfam; PF00623; RNA_pol_Rpb1_2_1.
DR Pfam; PF04983; RNA_pol_Rpb1_3_1.
DR Pfam; PF05000; RNA_pol_Rpb1_4_1.
DR Pfam; PF04998; RNA_pol_Rpb1_5_1.

DR SMART; SM00663; RPOLA N; 1.
 KM Transferrase; DNA-directed RNA polymerase; Transcription; Zinc;
 KW Zinc-finger; Nuclear protein. C2H2-TYPE (POTENTIAL).
 FT ZN_FING 64 80
 SQ SEQUENCE 1716 AA; 194191 MW; EBBE15BC33E60941 CRC64;

Alignment Scores:
 Pred. No.: 25 Length: 1716
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.20% Indels: 0
 DB: Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x RPA1_RAT (1-1716)

OY 386 AGTTCTCTCGTTGTTTCCTGCGC 409
 DB 1537 SerSerLeuValValSerLeuAla 1544

RESULT 16
 RPA1_HUMAN STANDARD; PRT; 1717 AA.
 ID RPA1_HUMAN
 AC O95602; Q9UEH0; Q9UEF7;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE DNA-directed RNA polymerase I largest subunit (EC 2.7.7.6) (RNA polymerase I 194 kDa subunit) (RPA194).
 GN POLR1A.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 ON NCBI_TaxID=9606;
 RX NCBL_TaxID=9606;
 RP SEQUENCE FROM N.A.
 RA Wang D., Stetler D.A.;
 RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 SEQUENCE OF 1623-1717 FROM N.A.
 RC TISSUE=Uterus;
 RA Koehler K., Beyer A., Mewes H.-W., Gassenhuber J., Wiemann S.;
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription of DNA into RNA using the four ribonucleoside triphosphates as substrates. RNA polymerase I is essentially used to transcribe ribosomal DNA units.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA} (N).
 CC -1- SUBUNIT: Each class of RNA polymerase is assembled from 9 to 14 different polypeptides. This subunit is the largest component of RNA polymerase I.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- MISCELLANEOUS: Three distinct zinc-containing RNA polymerases are found in eukaryotic nuclei: polymerase I for the ribosomal RNA precursor, polymerase II for the mRNA precursor, and polymerase III for 5S and tRNA genes.
 CC -1- SIMILARITY: Belongs to the RNA polymerase beta' chain family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U33460; AAC98959.1; -;
 DR EMBL; U33823; AAD09356.1; -;
 DR EMBL; U33824; AAD09356.1; JOINED.
 DR EMBL; AL117467; CAB55942.1; -;
 DR PIR; T17252; T17252.
 DR SWISS-2DPAGE; O95602; HUMAN.

DR Genew; HGNC:17264; POLR1A.
 DR GO; GO:0005736; C:DNA-directed RNA polymerase I complex; NAS.
 DR InterPro; IPR000722; RNA_pol_A.
 DR InterPro; IPR007080; RNA_pol_Rpb1_1.
 DR InterPro; IPR007066; RNA_pol_Rpb1_3.
 DR InterPro; IPR007083; RNA_pol_Rpb1_4.
 DR InterPro; IPR007081; RNA_pol_Rpb1_5.
 DR InterPro; IPR006592; RNA_pol_N.
 DR Pfam; PF04997; RNA_pol_Rpb1_1; 1.
 DR Pfam; PF00623; RNA_pol_Rpb1_2; 1.
 DR Pfam; PF04983; RNA_pol_Rpb1_3; 1.
 DR Pfam; PF05000; RNA_pol_Rpb1_4; 1.
 DR Pfam; PF04998; RNA_pol_Rpb1_5; 1.
 DR SMART; SM00663; RPOLA N; 1.
 KM Transferrase; DNA-directed RNA polymerase; Transcription; Zinc;
 KW Zinc-finger; Nuclear protein. C2H2-TYPE (POTENTIAL).
 FT ZN_FING 64 80
 FT CONFLICT 986 986 G -> A (IN REF. 1; AAD09356).
 FT CONFLICT 1372 1374 RAT -> TRI (IN REF. 1; AAD09356).
 SQ SEQUENCE 1717 AA; 194190 MW; 1143247B211F9BDB CRC64;

Alignment Scores:
 Pred. No.: 25 Length: 1717
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.20% Indels: 0
 DB: Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x RPA1_HUMAN (1-1717)

OY 386 AGTTCTCTCGTTGTTTCCTGCGC 409
 DB 1539 SerSerLeuValValSerLeuAla 1546

RESULT 17
 RPA1_MOUSE STANDARD; PRT; 1717 AA.
 ID RPA1_MOUSE
 AC O35134;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE DNA-directed RNA polymerase I largest subunit (EC 2.7.7.6) (RNA polymerase I 194 kDa subunit) (RPA194).
 GN POLR1A OR RPO1-4 OR RPA1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 ON NCBI_TaxID=10090;
 RX NCBL_TaxID=10090;
 RP SEQUENCE FROM N.A.
 RA MEDLINE=97379976; PubMed=9236775;
 RA Seither P., Coy J.F., Pouška A., Grunnt I.;
 RL Molecular cloning and characterization of the cDNA encoding the RT largest subunit of mouse RNA polymerase I.;
 RL Mol. Gen. Genet. 255:180-186(1997).
 CC -1- FUNCTION: DNA-dependent RNA polymerase catalyzes the transcription of DNA into RNA using the four ribonucleoside triphosphates as substrates. RNA polymerase I is essentially used to transcribe ribosomal DNA units.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + {RNA} (N).
 CC -1- SUBUNIT: Each class of RNA polymerase is assembled from 9 to 14 different polypeptides. This subunit is the largest component of RNA polymerase I.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- MISCELLANEOUS: Three distinct zinc-containing RNA polymerases are found in eukaryotic nuclei: polymerase I for the ribosomal RNA precursor, polymerase II for the mRNA precursor, and polymerase III for 5S and tRNA genes.
 CC -1- SIMILARITY: Belongs to the RNA polymerase beta' chain family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF000938; AAB6718.1; -.
 DR PIR; T13961; T13961.
 DR MCD; MG1:1096397; Rpol-1-4.
 DR InterPro; IPR000722; RNA_pol_A.
 DR InterPro; IPR007080; RNA_pol_Rpb1_1.
 DR InterPro; IPR007086; RNA_pol_Rpb1_3.
 DR InterPro; IPR007083; RNA_pol_Rpb1_4.
 DR InterPro; IPR007081; RNA_pol_Rpb1_5.
 DR InterPro; IPR006592; RNA_polA_N.
 DR Pfam; PF04997; RNA_pol_Rpb1_1.
 DR Pfam; PF00623; RNA_pol_Rpb1_2.
 DR Pfam; PF04983; RNA_pol_Rpb1_3.
 DR Pfam; PF05000; RNA_pol_Rpb1_4.
 DR Pfam; PF04998; RNA_pol_Rpb1_5.
 DR SMART; SM00663; RPOLA_N.1.
 DR Transferrase; DNA-directed RNA polymerase; Transcription; Zinc;
 DR Zinc-finger; Nuclear protein. CMH2-TYPE (POTENTIAL).
 FT ZN FING 64
 SQ SEQUENCE 1717 AA; 194079 MW; 77C3F23F27FD29BE CRC64;
 Alignment Scores:
 Pred. No.: 25 Length: 1717
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.20% Indels: 0
 DB: 1 Gaps: 0
 US-10-029-345A-108_COPY_532_2532 (1-2001) x RPA1_MOUSE (1-1717)
 Oy 386 AGTCTCGCTGTTGCTTCCCTGACC 409
 Db 1538 SerSerLeuValValSerLeuAla 1545
 RESULT 18
 POLN_SINDV STANDARD; PRT; 2512 AA.
 ID POLN_SINDV STANDARD; PRT; 2512 AA.
 AC P03317;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Nonstructural polyprotein (P270) [Contains: Nonstructural protein
 DE NSP1; Nonstructural NSP2; Nonstructural NSP3; Nonstructural NSP4].
 OS Sindbis virus (strain HRSF).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
 OC Alphavirus.
 OC NCBI_TaxID=11034;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=84148439; PubMed=6322438;
 RA Straus E.G., Rice C.M., Straus J.H.;
 RT "Complete nucleotide sequence of the genomic RNA of Sindbis virus";
 RL Virology 133:92-110(1984).
 RN [2]
 RP SEQUENCE OF 1-54 FROM N.A.
 RX MEDLINE=83268700; PubMed=6308269;
 RA Ou J.H., Straus E.G., Straus J.H.;
 RT "The 5'-terminal sequences of the genomic RNAs of several
 RL alphaviruses";
 RN [3]
 RP SEQUENCE OF 1429-2512 FROM N.A.
 RX MEDLINE=83299955; PubMed=6577423;
 RA Straus E.G., Rice C.M., Straus J.H.;
 RT "Sequence coding for the alphavirus nonstructural proteins is
 RT interrupted by an opal termination codon.";

RL Proc. Natl. Acad. Sci. U.S.A. 80:5271-5275(1983).
 RN [4]
 RP SEQUENCE OF 2431-2512 FROM N.A.
 RX MEDLINE=83039346; PubMed=6291034;
 RA Ou J.H., Rice C.M., Dalgarno L., Straus E.G., Straus J.H.;
 RT "Sequence studies of several alphavirus genomic RNAs in the region
 RL containing the start of the subgenomic RNA";
 RL Proc. Natl. Acad. Sci. U.S.A. 79:5235-5239(1982).
 CC -1- FUNCTION: NSP2 MAY BE INVOLVED IN RNA BINDING DURING REPLICATION.
 CC -1- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
 CC -1- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UGA OCCURS
 CC BETWEEN THE CODONS FOR 1896-TYR AND 1897-LEU.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; J02363; AAA96975.1; ALT_SEQ.
 DR PIR; A03917; MNWVS.
 DR MEROPS; C09.001; -.
 DR InterPro; IPR002589; A1pD.
 DR InterPro; IPR002620; peptidase_C9.
 DR InterPro; IPR001788; RNA_dep_RNAPol2.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR InterPro; IPR006066; Viral_helicase1.
 DR Pfam; PF01661; A1pD; 1.
 DR Pfam; PF01707; peptidase_C9; 1.
 DR Pfam; PF00978; RNA_dep_RNAPol2; 1.
 DR Pfam; PF01443; Viral_helicase1; 1.
 DR SMART; SM00506; A1pD; 1.
 DR PolyProtein; Nonstructural protein; RNA-binding; Helicase.
 FT CHAIN 1 540 NONSTRUCTURAL PROTEIN NSP1.
 FT CHAIN 541 1347 NONSTRUCTURAL PROTEIN NSP2.
 FT CHAIN 1348 1896 NONSTRUCTURAL PROTEIN NSP3.
 FT CHAIN 1897 2512 NONSTRUCTURAL PROTEIN NSP4.
 SQ SEQUENCE 2512 AA; 279546 MW; F3656FC8BB495726 CRC64;
 Alignment Scores:
 Pred. No.: 23 9 Length: 2512
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.20% Indels: 0
 DB: 1 Gaps: 0
 US-10-029-345A-108_COPY_532_2532 (1-2001) x POLN_SINDV (1-2512)
 Oy 1550 CCAGCTTCCTTTGGCCTTTCCA 1573
 Db 445 ProAlaSerPheSerAlaPhePro 452
 RESULT 19
 POLN_SINDV STANDARD; PRT; 2514 AA.
 ID POLN_SINDV STANDARD; PRT; 2514 AA.
 AC P27283;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Nonstructural polyprotein (P270) [Contains: Nonstructural protein
 DE NSP1; Nonstructural NSP2; Nonstructural NSP3; Nonstructural NSP4].
 OS Sindbis virus (subtype Ockelbo / strain Edabyn 82-5).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
 OC Alphavirus.
 OC NCBI_TaxID=31699;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91220725; PubMed=1673813;
 RA Shirako Y., Niklasson B., Dalrymple J.M., Straus E.G., Straus J.H.;

```

RL Mamm. Genome 6:746-748(1995).
[2]
RP MUTAGENESIS.
RC STRAIN=Sprague-Dawley, and Fischer 344/N;
RX MEDLINE=95148647; PubMed=7846077;
RA Katiuchi H., Watanabe M., Ushijima T., Toyota M., Imai K.,
RA Weisburger J.H., Sugimura T., Nagao M.;
RT "Specific 5'-GGGA-3'-->5'-GGA-3' mutation of the Apc gene in rat colon
RT tumors induced by 2-amino-1-methyl-6-pyrimidinazo(4,5-b)pyridine.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:910-914(1995).
CC -1- FUNCTION: Tumor suppressor. Promotes rapid degradation of CTNNB1
CC and participates in Wnt signaling. APC activity is correlated with
CC its phosphorylation state (by similarity).
CC -1- SUBUNIT: Forms homooligomers. Associates with catenins. Binds
CC axin (by similarity).
CC -1- PTM: Phosphorylated by GSK3B (by similarity).
CC -1- SIMILARITY: Contains 7 ARM repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcement/
CC or send an email to license@ebi.ac.uk).
CC -----
CC DR EMBL; D38629; BAA07609.1; -.
CC DR HSSP; Q02248; 3BCT.
CC DR InterPro; IPR008938; ARM.
CC DR InterPro; IPR00225; Armadillo.
CC DR Pfam; PF00514; Armadillo_seg; 4.
CC DR SMART; SM00185; ARM; 5.
CC DR PROSITE; PS50176; ARM_REPEAT; 1.
CC KW Wnt signaling pathway; Anti-oncogene; Phosphorylation; Coiled coil;
CC Repeat.
CC
CC FT DOMAIN 1 728 LEU-RICH.
CC FT DOMAIN 1 62 COILED COIL (POTENTIAL).
CC FT DOMAIN 125 260 COILED COIL (POTENTIAL).
CC FT REPEAT 451 493 ARM 1.
CC FT REPEAT 503 545 ARM 2.
CC FT REPEAT 546 589 ARM 3.
CC FT REPEAT 590 636 ARM 4.
CC FT REPEAT 637 681 ARM 5.
CC FT REPEAT 682 723 ARM 6.
CC FT REPEAT 724 765 ARM 7.
CC FT DOMAIN 739 2831 SER-RICH.
CC FT DOMAIN 1130 1155 ASP/GLU-RICH (ACIDIC).
CC FT DOMAIN 1556 1575 ASP/GLU-RICH (ACIDIC).
CC FT DOMAIN 1864 1891 HIGHLY CHARGED.
CC FT DOMAIN 523 523 C-SH: IN AN IQ-INDUCED COLON TUMOR.
CC
CC SQ SEQUENCE 2842 AA; 310530 MW; 3CB82BA8A34EBF47 CRC64;
CC
CC Alignment Scores:
CC Pred. No.: 23.6 Length: 2842
CC Score: 8.00 Matches: 8
CC Percent Similarity: 100.00% Conservative: 0
CC Best Local Similarity: 100.00% Mismatches: 0
CC Query Match: 1.24% Indels: 0
CC DB: 1 Gaps: 0
CC
CC US-10-029-345A-108_COPY_532_2532 (1-2001) x APC_RAT (1-2842)
CC
CC QY 959 TTCTCAGTCGACGACGCTTGACT 936
CC |||||||||||||||||||||||||
CC DB 1268 PheserArgCysserSerLeuser 1275
CC
CC RESULT 21
CC APC_HUMAN STANDARD; PRT; 2843 AA.
CC AC P25054; Q15162; Q15163;
CC DT 01-MAY-1992 (Rel. 22, Created)
CC DT 01-MAY-1992 (Rel. 22, Last sequence update)
CC DT 15-MAR-2004 (Rel. 43, Last annotation update)

```

DE Adenomatous polyposis coli protein (APC protein).
 CN APC OR DP2.5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 CX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=91335210; PubMed=1651562;
 RA Kinzler K.W., Nilbert M.C., Su L.-K., Vogelstein B., Bryan T.M.,
 RA Levy D.B., Smith K.J., Preisinger A.C., Hedge P., McKechnie D.,
 RA Finnian R., Markham A., Groffen J., Boguski M.S., Altschul S.F.,
 RA Horii A., Ando H., Miyoshi Y., Miki Y., Nishisho I., Nakamura Y.,
 RT "Identification of FAP locus genes from chromosome 5q21.";
 RL Science 253:661-665 (1991).
 RN [2]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RC TISSUE=Fetal brain;
 RX MEDLINE=91330307; PubMed=1678319;
 RA Joslyn G., Carlson M., Thliveris A., Albertsen H., Gelbert L.,
 RA Samowitz W., Groden J., Stevens J., Spirio L., Robertson M.,
 RA Sargeant L., Krapcho K., Wolff E., Burt R., Hughes J.P.,
 RA Warrington J., McPherson J.D., Masmuth J., Le Faslter D.,
 RA Abderrahim H., Cohen D., Leppert M., White R.,
 RT "Identification of deletion mutations and three new genes at the
 RT familial polyposis locus.";
 RL Cell 66:601-613 (1991).
 RN [3]
 RP ASSOCIATION WITH CATENINS.
 RX MEDLINE=94082295; PubMed=8259519;
 RA Su L.-K., Vogelstein B., Kinzler K.W.,
 RT "Association of the APC tumor suppressor protein with catenins.";
 RL Science 262:1734-1737 (1993).
 RN [4]
 RP DISEASE.
 RX MEDLINE=95174843; PubMed=7661930;
 RA Hamilton S.R., Liu B., Parsons R.E., Papadopoulos N., Jen J.,
 RA Powell S.M., Krush A.J., Berk T., Cohen Z., Iacu B., Burger P.C.,
 RA Wood P.A., Tadi F., Booker S.V., Petersen G.M., Offerhaus G.J.A.,
 RA Teremette A.C., Giardiello F.M., Vogelstein B., Kinzler K.W.,
 RT "The molecular basis of Turcot's syndrome.";
 RL New Engl. J. Med. 332:839-847 (1995).
 RN [5]
 RP DISEASE.
 RX MEDLINE=97094176; PubMed=8940264;
 RA Eccles D.M., van der Luij R.B., Breukel C., Bullman H., Bunyan D.,
 RA Fisher A., Barber J., du Boulay C., Primrose J., Burn J., Fodde R.,
 RT "Hereditary desmoid disease due to a frameshift mutation at codon 1924
 RL of the APC gene.";
 RL Am. J. Hum. Genet. 59:1193-1201 (1996).
 RN [6]
 RP DISEASE.
 RX MEDLINE=20243021; PubMed=10782927;
 RA Couture J., Mitri A., Lagace R., Smits R., Berk T., Bouchard H.-L.,
 RA Fodde R., Altman B., Bapat B.,
 RT "A germline mutation at the extreme 3'-prime end of the APC gene
 RT results in a severe desmoid phenotype and is associated with
 RT overexpression of beta-catenin in the desmoid tumor.";
 RL Clin. Genet. 57:205-212 (2000).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 2-55.
 RA MEDLINE=20384842; PubMed=10926498;
 RX Day C.L., Alber T.,
 RT "Crystal structure of the amino-terminal coiled-coil domain of the
 RT APC tumor suppressor.";
 RL J. Mol. Biol. 301:147-156 (2000).
 RN [8]
 RP X-RAY CRYSTALLOGRAPHY (3.1 ANGSTROMS) OF 1021-1035 IN COMPLEX WITH
 RP CTNNB1.
 RX MEDLINE=21564054; PubMed=11707392;
 RA Eklof Spink K., Fridman S.G., Weis W.I.,
 RT "Molecular mechanisms of beta-catenin recognition by adenomatous
 RT polyposis coli revealed by the structure of an APC-beta-catenin
 RT complex.";
 RL EMBO J. 20:6203-6212 (2001).
 RN [9]
 RP X-RAY CRYSTALLOGRAPHY (1.90 ANGSTROMS) OF 2034-2049 IN COMPLEX WITH
 RP AXIN.
 RX MEDLINE=20271867; PubMed=10811618;
 RA Spink K.E., Polakis P., Weis W.I.,
 RT "Structural basis of the axin-adenomatous polyposis coli
 RT interaction.";
 RL EMBO J. 19:2270-2279 (2000).
 RN [10]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=94154728; PubMed=8111410;
 RA Nagase H., Nakamura Y.,
 RT "Mutations of the APC (adenomatous polyposis coli) gene.";
 RL Hum. Mutat. 2:425-434 (1993).
 RN [11]
 RP VARIANTS FAP.
 RX MEDLINE=91335211; PubMed=1651563;
 RA Nishisho I., Nakamura Y., Miyoshi Y., Miki Y., Ando H., Horii A.,
 RA Koyama K., Utsunomiya J., Baba S., Hedge P., Markham A., Krush A.J.,
 RA Petersen G.M., Hamilton S.R., Nilbert M.C., Levy D.B., Bryan T.M.,
 RA Preisinger A.C., Smith K.J., Su L.-K., Kinzler K.W., Vogelstein B.,
 RT "Mutations of chromosome 5q21 genes in FAP and colorectal cancer
 RT patients.";
 RL Science 253:665-669 (1991).
 RN [12]
 RP VARIANTS FAP.
 RX MEDLINE=93265030; PubMed=1338904;
 RA Miyoshi Y., Nagase H., Ando H., Ichii S., Nakatsuru S., Aoki T.,
 RA Miki Y., Mori T., Nakamura Y.,
 RT "Somatic mutations of the APC gene in colorectal tumors: mutation
 RT cluster region in the APC gene.";
 RL Hum. Mol. Genet. 1:229-233 (1992).
 RN [13]
 RP VARIANTS FAP.
 RX MEDLINE=93244793; PubMed=1338691;
 RA Nakatsuru S., Yanagisawa A., Ichii S., Tahara E., Kato Y.,
 RA Nakamura Y., Horii A.,
 RT "Somatic mutation of the APC gene in gastric cancer: frequent
 RT mutations in very well differentiated adenocarcinoma and signet-ring
 RT cell carcinoma.";
 RL Hum. Mol. Genet. 1:559-563 (1992).
 RN [14]
 RP VARIANTS FAP TRP-1348, AND VARIANTS ASP-1118; MET-1292; VAL-1304 AND
 RP SER-2502.
 RX MEDLINE=93250848; PubMed=1338764;
 RA Nagase H., Miyoshi Y., Horii A., Aoki T., Petersen G.M.,
 RA Vogelstein B., Maher E., Ogawa M., Maruyama M., Utsunomiya J.,
 RA Baba S., Nakamura Y.,
 RT "Screening for germ-line mutations in familial adenomatous polyposis
 RT patients: 61 new patients and a summary of 150 unrelated patients.";
 RL Hum. Mutat. 1:467-473 (1992).
 RN [15]
 RP VARIANTS FAP TRP-99.
 RC TISSUE=Peripheral blood lymphocytes;
 RX MEDLINE=95134544; PubMed=7833149;
 RA Dobbie Z., Spycher M., Huertlman R., Ammann R., Ammann T., Roth J.,
 RA Mueller A., Mueller H., Scott R.J.,
 RT "Mutational analysis of the first 14 exons of the adenomatous
 RT polyposis coli (APC) gene.";
 RL Eur. J. Cancer 30A:1709-1713 (1994).
 RN [16]
 RP VARIANTS FAP GUY-722.
 RX MEDLINE=95135430; PubMed=7833931;
 RA Stella A., Montero M., Reeta N., Marchese C., Susca F., Gentile M.,
 RA Romio L., Pilla S., Prete F., Marenzi C., Guanti G.,
 RT "Four novel mutations of the APC (adenomatous polyposis coli) gene in
 RT FAP patients.";
 RL Hum. Mol. Genet. 3:1687-1688 (1994).
 RN [17]
 RP ERRATUM.
 RA Stella A., Montero M., Reeta N., Marchese C., Susca F., Gentile M.,

RA Romio L., Pilla S., Prete F., Mareni C., Guanti G.;
 RL Hum. Mol. Genet. 3:1918-1918(1994).
 RN [18]
 RP VARIANT PAP ILE-171.
 RX MEDLINE=97144176; PubMed=8990002;
 RA van der Luit R.B., Meera Khan P., Vaseen H.F.A., Tops C.M.J.,
 RA van Leeuwen-Cornelisse I.S.J., Wijnen J.T., van der Klift H.M.,
 RA Plug R.J., Griffioen G., Fodde R.,
 RT "Molecular analysis of the APC gene in 105 Dutch kindreds with
 RT familial adenomatous polyposis: 67 germline mutations identified by
 RT DGGE, PTT, and southern analysis."
 RL Hum. Mutat. 9:7-16(1997).
 RN [19]
 RP VARIANTS COLORRECTAL CARCINOMA THR-880; ILE-890 AND VAL-1508.
 RX MEDLINE=98080146; PubMed=9419979;
 RA Miyaki M., Nishio J., Konishi M., Kikuchi-Yanoshita R., Tanaka K.,
 RA Muraoka M., Nagato M., Chong J.-M., Koike M., Terada T., Kawahara Y.,
 RA Fukuyama A., Tomiyama J., Chuganji Y., Momoi M., Urunomiya J.,
 RT "Drastic genetic instability of tumors and normal tissues in Turcot
 RT syndrome."
 RL Oncogene 15:2877-2881(1997).
 RN [20]
 RP VARIANT LYS-1307.
 RX MEDLINE=98400248; PubMed=9731522;
 RA Rederom M., Nathanson K.L., Yuan Z.Q., Neuhausen S.L., Saragopan J.,
 RA Wong N., Yang D., Nafa D., Abrahamson J., Ozcelik H.,
 RA Antin-Ozerkis D., Andrulis I., Daly M., Pinsky L., Schrag D.,
 RA Gallinger S., Kaback M., King M.-C., Woodage T., Brody L.C.,
 RA Godwin A., Warner E., Weber B., Foulkes W., Offit K.,
 RT "The APC I1307K allele and breast cancer risk."
 RL Nat. Genet. 20:13-14(1998).
 RN [21]
 RP VARIANTS LYS-1307 AND GLN-1317.
 RX TISSUE=Peripheral blood;
 RC MEDLINE=98393712; PubMed=9724771;
 RA Praying I.M., Beck N.E., Ilyas M., Dove-Edwin I., Goodman P.,
 RA Pack K., Bell J.A., Williams C.B., Hodgson S.V., Thomas H.J.W.,
 RA Talbot I.C., Bodmer W.F., Tomlinson I.P.M.,
 RT "The APC variants I1307K and E1317Q are associated with colorectal
 RT tumors, but not always with a family history."
 RL Proc. Natl. Acad. Sci. U.S.A. 95:10722-10727(1998).
 RN [22]
 RP VARIANT LYS-1307.
 RX MEDLINE=98400259; PubMed=9731533;
 RA Woodage T., King S.W., Wacholder S., Hartge P., Struwing J.P.,
 RA McAdams M., Iaken S.J., Tucker M.A., Brody L.C.,
 RT "The APC I1307K allele and cancer risk in a community-based study of
 RT Ashkenazi Jews."
 RL Nat. Genet. 20:62-65(1998).

Alignment Scores:

Pred. No.:	23.6	Length:	2843
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.24%	Indels:	0
DB:	1	Gaps:	0

US-10-029-345a-108_COPY_532_2532 (1-2001) x APC_HUMAN (1-2843)

QY 959 TTCTCCAGGTGCAGCGCTTGACT 936

DB 1271 PheSerArgCysSerSerLeuSer 1278

RESULT 22

APC_MOUSE STANDARD; PRT; 2845 AA.

AC 061315; 062044;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Adenomatous polyposis coli protein (APC protein) (mAPC).
 GN APC.

OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND VARIANTS.
 RX STRAIN=C57BL/6J, and CAST/EI; TISSUE=Brain;
 RC MEDLINE=92263101; PubMed=1350108;
 RA Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,
 RA Luongo C., Gould K.A., Dove W.F.,
 RT "Multiple intestinal neoplasia caused by a mutation in the murine
 RT homolog of the APC gene."
 RL Science 256:668-670(1992).
 RN [2]
 RP ERRATUM.
 RA Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,
 RA Luongo C., Gould K.A., Dove W.F.,
 RL Science 256:1114-1114(1992).
 RN [3]
 RP SEQUENCE OF 1-45 FROM N.A.
 RC STRAIN=BALB/c; TISSUE=Liver;
 RA Dicker F., Lambertz S., Reilmair A., Ballhausen W.G.,
 RT "The murine APC gene: alternative splicing of 5' untranslated
 RT region segments."
 RL Submitted (OCT-1993) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP ALTERNATIVE SPLICING.
 RX MEDLINE=94061824; PubMed=8242607;
 RA Oshima M., Sugiyama H., Kitagawa K., Taketo M.,
 RA "APC gene messenger RNA: novel isoforms that lack exon 7."
 RL Cancer Res. 53:5589-5591(1993).
 CC -1- FUNCTION: Tumor suppressor. Promotes rapid degradation of CTNNB1
 CC and participates in Wnt signaling. APC activity is correlated with
 CC its phosphorylation state (By similarity).
 CC -1- SUBUNIT: Forms homooligomers. Associates with catenine. Binds
 CC axin (By similarity).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=4;
 CC Name=1;
 CC IsoId=Q61315-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q61315-2; Sequence=VSP_004116;
 CC Name=3;
 CC IsoId=Q61315-3; Sequence=VSP_004117;
 CC Name=4;
 CC IsoId=Q61315-4; Sequence=VSP_004116, VSP_004117;
 CC -1- TISSUE SPECIFICITY: Expressed in liver, spleen, kidney, heart,
 CC lung, brain, stomach, intestine, testis and ovary.
 CC -1- PTM: Phosphorylated by GSK3B (By similarity).
 CC -1- SIMILARITY: Contains 7 ARM repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; M88127; AAB59632.1; -
 DR EMBL; U02937; AAA03443.1; -
 DR FIR; 149505; 149505.
 DR HSSP; 002248; 3BCT.
 DR MGD; MGI:88039; APC.
 DR GO; GO:0005737; Cytoplasm; IDA.
 DR GO; GO:0005634; Cytoleus; IDA.
 DR GO; GO:0008013; F-beta-catenin binding; IDA.
 DR GO; GO:0009952; Planterior/posterior pattern formation; IMP.
 DR GO; GO:0009798; Paxis specification; IMP.
 DR GO; GO:0009953; P:dorsal/ventral pattern formation; IMP.
 DR GO; GO:0016055; P:Wnt receptor signaling pathway; IDA.
 DR InterPro; IPR008938; ARM.
 DR InterPro; IPR000225; Armadillo.

DR Pfam; PF00514; Armadillo_seg; 4.
 DR SMART; SM00185; ARM; 5.
 DR PROSITE; PS00176; ARM_REPEAT; 1.
 KW mt signaling pathway; Anti-oncogene; Phosphorylation;
 KW Alternative splicing; Repeat; Coiled coil.
 FT DOMAIN 1 61 COILED COIL (POTENTIAL).
 FT DOMAIN 125 245 COILED COIL (POTENTIAL).
 FT DOMAIN 1 728 COILED COIL (POTENTIAL).
 FT REPEAT 451 493 ARM 1.
 FT REPEAT 503 545 ARM 2.
 FT REPEAT 546 589 ARM 3.
 FT REPEAT 590 636 ARM 4.
 FT REPEAT 637 681 ARM 5.
 FT REPEAT 682 723 ARM 6.
 FT REPEAT 724 765 ARM 7.
 FT DOMAIN 739 2834 SER-RICH.
 FT DOMAIN 1130 1156 ASP/GLU-RICH (ACIDIC).
 FT DOMAIN 1556 1575 ASP/GLU-RICH (ACIDIC).
 FT DOMAIN 1864 1891 HIGHLY CHARGED.
 FT VARSPLIC 243 276 Missing (in isoform 2 and isoform 4).
 FT VARSPLIC 310 410 Missing (in isoform 3 and isoform 4).
 FT VARSPLIC 120 120 /FTid=VSP_004117.
 FT VARIANT 493 493 T -> A (IN STRAIN CAST/ET).
 FT VARIANT 797 797 V -> I (IN STRAIN CAST/ET).
 FT VARIANT 797 797 Y -> F (IN STRAIN CAST/ET).
 FT VARIANT 1330 1330 A -> T (IN STRAIN CAST/ET).
 FT VARIANT 1618 1618 G -> S (IN STRAIN CAST/ET).
 FT VARIANT 2294 2294 A -> A (IN STRAIN CAST/ET).
 FT VARIANT 2496 2496 H -> Q (IN STRAIN CAST/ET).
 FT VARIANT 2523 2523 T -> A (IN STRAIN CAST/ET).
 FT VARIANT 2813 2813 T -> S (IN STRAIN CAST/ET).
 SQ SEQUENCE 2845 AA; 311086 MW; 145CA73CF570A499 CRC64;

Alignment Scores:
 Pred. No.: 23.6 Length: 2845
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.24% Indels: 0
 DB: 1 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x APC_MOUSE (1-2845)

OY 959 TTCTCGAGTGCACGACCTTGAGT 936
 DB 1269 PheserargCysSerSerLeuSer 1276

RESULT 23
 ABGA_CLOLO STANDARD; PRT; 473 AA.
 ID ABGA_CLOLO
 AC 046130;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 6-phospho-beta-glucosidase (EC 3.2.1.86).
 GN ABGA.
 OS Clostridium longisporum.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 NCBI_TaxID=1523;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=B6405;
 RA Brown G.D.; Thomson J.A.;
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: 6-phospho-beta-D-glucoside-(1,4)-D-glucose +
 H(2)O = D-glucose 6-phosphate + D-glucose.
 CC -1- SIMILARITY: Belongs to family 1 of glycosyl hydrolases.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; LA9336; AAC05714.1; -.
 DR HSSP; P11546; 1PBG.
 DR InterPro; IPR001360; Glyco_hydro_1.
 DR Pfam; PF00232; Glyco_hydro_1; 1.
 DR PRINTS; PR00131; GLYHYDRLASE1.
 DR ProDom; PD000650; Glyco_hydro_1; 1.
 DR PROSITE; PS00572; GLYCOSYL_HYDROL_F1_1; 1.
 DR PROSITE; PS00653; GLYCOSYL_HYDROL_F1_2; 1.
 KW Hydrolase; Glycosylase.
 FT ACT SITE 174 174 PROTON DONOR (POTENTIAL).
 FT ACT SITE 366 366 NUCLEOPHILE (BY SIMILARITY).
 SQ SEQUENCE 473 AA; 54413 MW; 9862BA0023903ECD CRC64;

Alignment Scores:
 Pred. No.: 314 Length: 473
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.09% Indels: 0
 DB: 1 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x ABGA_CLOLO (1-473)

OY 1791 ATGACCTTGTCCTCCGCAAGT 1771
 DB 420 IleAspLeuValSerAlaSer 426

RESULT 24
 ATPB_PTEES STANDARD; PRT; 473 AA.
 ID ATPB_PTEES
 AC 003080;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE ATP synthase beta chain (EC 3.6.3.14) (Fragment).
 GN ATPB.
 OS Pteridium esculentum.
 OC Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Moniliformopses; Filicophyta; Filicopsida; Filicales;
 OC Dennstaedtiaceae; Pteridium.
 NCBI_TaxID=32102;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Wolf P.G.;
 RL "Evaluation of atpB nucleotide sequences for phylogenetic studies of
 RT ferns and other pteridophytes";
 RL Am. J. Bot. 84:1429-1440(1997).
 CC -1- FUNCTION: Produces ATP from ADP in the presence of a proton
 CC gradient across the membrane. The beta chain is the catalytic
 CC subunit.
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (In) = ADP + phosphate +
 H(+) (Out).
 CC -1- SUBUNIT: F-type ATPases have 2 components, CF(1) - the catalytic
 CC core - and CF(0) - the membrane proton channel. CF(1) has five
 CC subunits: alpha(3), beta(3), gamma(1), delta(1), epsilon(1). CF(0)
 CC has three main subunits: a, b and c.
 CC -1- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.
 CC -1- SIMILARITY: Belongs to the ATPase alpha/beta chains family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```
DR EMBL: U93834; AAB51742.1; -.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR005722; ATP_synthf1_beta.
DR InterPro: IPR00793; ATPase_a/b_C.
DR InterPro: IPR00194; ATPase_a/b_centre.
DR InterPro: IPR004100; ATPase_a/bN.
DR InterPro: IPR009005; F1_ATPase_a/bN.
DR Pfam: PF00006; ATP-synt_ab: 1.
DR Pfam: PF00306; ATP-synt_ab_C: 1.
DR Pfam: PF02874; ATP-synt_ab_N: 1.
DR SMART: SM00382; AAA; 1.
DR TIGRPFAMs: TIGR01039; atpD; 1.
DR PROSITE: PS00152; ATPase_ALPHA_BETA; 1.
DR ATP synthase; Chloroplast; Thylakoid; Membrane; CF(1);
KW Hydrolyase; ATP-binding; Hydrogen ion transport.
FT NON_TER 1
FT NP_BIND 172 179 ATP (POTENTIAL).
FT NON_TER 473
FT NON_TER 473
SQ SEQUENCE 473 AA; 50434 MW; 5F305968B8B90029 CRC64;

Alignment Scores:
Pred. No.: 314 Length: 473
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.05% Indels: 0
DB: 1 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x ATPB_PTEES (1-473)
OY 193 GTGTTAATTACAGAGCATC 213
DB 180 ValLeu1eHnrgLueu1e 186

RESULT 25
ASCB_ECOLI STANDARD; PRT; 474 AA.
ID AC P24240; P78104; Q59375;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 6-phospho-beta-glucosidase ascb (EC 3.2.1.86).
GN ASCB OR B2716.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=92334140; PubMed=1630307;
RA Hall B.G., Xu L.;
RT "Nucleotide sequence, function, activation, and evolution of the
RT cryptic asc operon of Escherichia coli K12.";
RL Mol. Biol. Evol. 9:688-706(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glaeser J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.O.,
RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
CC -1- FUNCTION: CAN HYDROLYSE SALICIN, CELLOBIOSE, AND PROBABLY
CC ARABUTIN.
CC -1- CATALYTIC ACTIVITY: 6-phospho-beta-D-glucoside-(1,4)-D-glucose +
CC H2O = D-glucose 6-phosphate + D-glucose.
CC -1- SIMILARITY: Belongs to family 1 of glycosyl hydrolases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M73326; AAA16430.1; -.
DR EMBL: U29579; AAA69226.1; ALT_INIT.
DR EMBL: AE000355; AAC75758.1; -.
DR PIR: H65051; H65051.
DR HSSP: P11546; IPBG.
DR EcoGene: EG10085; ascb.
DR InterPro: IPR001360; Glyco_hydro_1.
DR Pfam: PF00232; Glyco_hydro_1; 1.
DR PRINTS: PR00131; GLHYDRLASE1.
DR PRODOM: PD000650; Glyco_hydro_1; 1.
DR PROSITE: PS00572; GLYCOSYL_HYDROL_F1_1; 1.
DR PROSITE: PS00653; GLYCOSYL_HYDROL_F1_2; 1.
KW Hydrolyase; Glycosidase; Complete proteome.
FT ACT_SITE 180 180 PROTON DONOR (POTENTIAL).
FT ACT_SITE 372 372 NUCLEOPHILE (BY SIMILARITY).
FT CONFLICT 405 406 EA -> GT (IN REF. 1).
FT CONFLICT 428 428 S -> C (IN REF. 1).
FT CONFLICT 455 456 RK -> HR (IN REF. 1).
SQ SEQUENCE 474 AA; 53935 MW; 02ACE6B8BF211011 CRC64;

Alignment Scores:
Pred. No.: 314 Length: 474
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.09% Indels: 0
DB: 1 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x ASCB_ECOLI (1-474)
OY 1791 ATGACTGTGTCGCGCAGT 1771
DB 422 l1eapleuvalserlaser 428

RESULT 26
PERI_MOUSE STANDARD; PRT; 475 AA.
ID AC P15311; Q35688; Q35689;
DT 01-APR-1990 (Rel. 14, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Peripherin.
GN PRPH OR PRPH1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RX MEDLINE=93192779; PubMed=1294287;
RA Karpoz V., Landon F., Djabali K., Gros F., Portier M.M.;
RT "Structure of the mouse gene encoding peripherin: a neuronal
RT intermediate filament protein.";
RL Biol. Cell 76:43-48(1992).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM SG).
RC STRAIN=FVB/N; TISSUE=Colon;
RX MEDLINE=22388257; PubMed=12477932;
RA Strassberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altshul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Hsieh P.,
RA Diatchenko L., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uediri T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
```

```

Alignment Scores:
Pred. No.: 314
Score: 7.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 1.05%
DB: 1

US-10-029-345A-108_COPY_532_2532 (1-2-2001) x PERI_MOUSE (1-475)
Qy 758 CGCGCCACATCGCTATCGCCT 778
Db 30 ProProteinServerPro 36

RESULT 27
SIM1 YEAST
ID SIM1 YEAST STANDARD: PRT: 475 AA.
AC P40472;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE SIM1 protein precursor.
GN SIM1 OR PB3 OR YII123W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
OC NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SS288C / AB972;
RX MEDLINE=97313266; PubMed=9168870;
RA Churcher C.M., Bowman S., Badcock K., Bankier A., Brown D.,
RA Chillingworth T., Connor R., Devlin K., Gentles S., Hamlin N.,
RA Harris D.E., Hortsnell T., Hunt S., Jagals K., Jones M., Lye G.,
RA Moulé S., Odell C., Pearson D., Rajandream M.A., Rice P., Rowley N.,
RA Skelton J., Smith V., Walsh S., Whitehead S., Barrett B.G.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome IX.";
RL Nature 387:84-87(1997).
RN [2]
RP IDENTIFICATION.
RX MEDLINE=96155611; PubMed=8574583;
RA Dahmann C., Diffley J.F.X., Naemnyth K.A.;
RT "8-phase-promoting cyclin-dependent kinases prevent re-replication by
RT inhibiting the transition of replication origins to a pre-replicative
RT state.";
RL Curr. Biol. 5:1257-1269(1995).
CC -! FUNCTION: REQUIRED FOR THE MAINTENANCE OF THE CLAS KINASE
CC ACTIVITY.
CC -! SIMILARITY: BELONGS TO THE SUN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; 246833; CAA86669.1; -.
DR DR PIR; S49886; S49886.
DR GERMOnline; 139658; -.
DR SGP; S0001385; STM1.
DR GO; GO:0002927; C:cell wall (sensu Fungi); IDA.
DR GO; GO:0000226; P:microtubule cytoskeleton organization and b. .; IGI.
DR InterPro; IPR005556; SUN.
DR Pfam; PF03856; SUN; 1.
KW Signal.
FT SIGNAL. 1 19 POTENTIAL.
FT CHAIN 20 475 SIM1 PROTEIN.
FT DOMAIN 58 112 ALA-RICH.
FT DOMAIN 80 203 SER-RICH.
FT CARBOHYD 422 423

```

Alignment Scores:

Pred. No.: 314 Length: 475
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.09% Indels: 0
 DB: 1 Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x SIM1_YEAST (1-475)

QY 1058 CCTGCTGCTTGGAGTAGCA 1038
 |||||

Db 57 ProAlaAlaSerGIuValAla 63

RESULT 28

LXC2_PHOLE STANDARD; PRT; 478 AA.

AC P29236;

DT 01-DEC-1992 (Rel. 24, Last sequence update)

DT 01-JUN-1994 (Rel. 29, Last annotation update)

DE Acyl-CoA reductase (EC 1.2.1.50).

GN LUXC.

OC Bacteri; Photobacterium leiognathi.

OC Vibrionaceae; Photobacterium.

OC NCBI_TaxID=658;

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 25521;

RX MEDLINE=92007870; PubMed=1915359;

RA Lee C.Y., Saitter R.B., Melsgen E.A.;

RT "The lux genes of the luminous bacterial symbiont, Photobacterium

leiognathi, of the porphyrio. Nucleotide sequence, difference in gene

organization, and high expression in mutant Escherichia coli.";

RL Bur. J. Biochem. 201:161-167(1991).

CC -1- FUNCTION: LUXC IS THE FATTY ACID REDUCTASE ENZYME RESPONSIBLE FOR

SYNTHESIS OF THE ALDEHYDE SUBSTRATE FOR THE LUMINESCENT REACTION

CATALYZED BY LUCIFERASE.

CC -1- CATALYTIC ACTIVITY: A long-chain aldehyde + CoA + NADP(+) = a

long-chain acyl-CoA + NADPH.

CC -1- PATHWAY: Bioluminescent fatty acid reduction system; third step.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

Alignment Scores:

Pred. No.: 313 Length: 478
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.05% Indels: 0
 DB: 1 Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x LXC2_PHOLE (1-478)

QY 355 TCTGTTACCTGCTTGCAGGT 375
 |||||

Db 145 SerValHisLeuValAlaGly 151

RESULT 29

TR2_HUMAN STANDARD; PRT; 483 AA.

AC P13056;

DT 01-JAN-1990 (Rel. 13, Created)

DT 01-JAN-1990 (Rel. 13, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Orphan nuclear receptor TR2.

GN NR2C1 OR TR2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OC NCBI_TaxID=9606;

RP SEQUENCE FROM N.A.

RX MEDLINE=8833993; PubMed=3421977;

RA Chang C., Kokontis J.;

RT "Identification of a new member of the steroid receptor super-family

by cloning and sequence analysis.";

RL Biochem. Biophys. Res. Commun. 155:971-977(1988).

CC -1- FUNCTION: Orphan nuclear receptor.

CC -1- SUBCELLULAR LOCATION: Nuclear (potential).

CC -1- SIMILARITY: Belongs to the nuclear hormone receptor family. NR2

subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

CC EMBL; M21965; AAA3650.1; -.

DR PIR; A31521; A31521.

DR HSSP; P19793; 2NLL.

DR TRANSFAC; T01662; -.

DR Genew; HGNC:7971; NR2C1.

DR MIM; 601529; -.

DR GO; GO:0003677; F:DNA binding; TAS.

DR GO; GO:0004872; F:steroid hormone receptor activity; TAS.

DR GO; GO:0003707; F:steroid hormone receptor activity; TAS.

DR InterPro; IPR000536; Hormone_rec_1lg.

DR InterPro; IPR001723; Sterhmr_receptor.

DR InterPro; IPR008946; Str_ncl_receptor.

DR InterPro; IPR001628; Znf_C4steroid.

DR Pfam; PF00104; hormone_rec; 1.

DR Pfam; PF00105; zfc4; 1.

DR PRINTS; PR00398; STRDHOMONER.

DR PRINTS; PR00047; STROIDPRNGER.

DR ProDom; PD000035; Znf_C4; 1.

DR PROSITE; PS00031; NUCLEAR RECEPTOR; 1.

KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;

KW Zinc-finger.

FT DNA_BIND 113 178 NUCLEAR RECEPTOR-TYPE.

FT ZN_FING 113 133 C4-TYPE.

FT ZN_FING 149 173 C4-TYPE.

SEQUENCE 483 AA; 53042 MW; 293E21E317E05814 CRC64;

Alignment Scores:

Pred. No.: 313 Length: 483
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.09% Indels: 0
 DB: 1 Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x TR2_HUMAN (1-483)

QY 487 GGATGGCCCAATGTGGCA 467
 |||||

Db 437 GlyLeuAlaGlnCysTyrIle 443

```

DT 01-NOV-1990 (Rel. 16, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glycoprotein polypeptide precursor [Contains: Glycoprotein G1,
GN Glycoprotein G2].
OS Mopeia virus.
OS Viruses; ssRNA negative-strand viruses; Arenaviridae; Arenavirus.
OX NCBI_TaxId=11623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=800150;
RC MEDLINE=91111973; PubMed=1989384;
RC Wilson S.M., Cleeg J.C.S.'
"Sequence analysis of the S RNA of the African arenavirus Mopeia: an
RT unusual secondary structure feature in the intergenic region.";
RT Virology 180:543-552(1991).
CC -1 SIMILARITY: Belongs to the arenaviruses GPC protein family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; M38879; AAC08700.1; -
DR PIR; A38546; VGXPWV.
DR InterPro; IPR001535; Arena_glycoprot.
DR Pfam; PF00798; Arena_glycoprot; 1.
DR Polyprotein; Glycoprotein; Envelope protein.
FT CHAIN 1 255
FT FT 256 489 GLYCOPROTEIN G1.
FT CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 88 88 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 98 98 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 108 108 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 118 118 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 222 222 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 363 363 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 371 371 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 388 388 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 393 393 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 489 AA; 55987 MW; B188BBF327E170B1 CRC64;
Alignment Scores:
Pred. No.: 312
Score: 7.00 Length: 489
Percent Similarity: 100.00% Matches: 7
Best Local Similarity: 100.00% Conservative: 0
Query Match: 1.05% Mismatches: 0
DB: 1 Indels: 0
Gaps: 0
US-10-029-345A-108_COPY_532_2532 (1-2001) x VELV_M0PE1 (1-489)
QY 971 CTGTCCTGCTCTCTCAGAGC 991
Db 240 LeuSerLeuLeuSerGlnGln 246
RESULT 32
Y1J0 YEAST
ID Y1J0 YEAST STANDARD; PRT; 491 AA.
AC P40499;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical 56.3 kDa protein in SG41-KIR7 intergenic region.
GN Y1J090W.
OS Saccharomyces cerevisiae (Baker's yeast).
OS Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
NCBI_TaxId=4932;

```

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288C / AB972;
 RX MEDLINE=97313266; PubMed=9169870;
 RA Churcher C.M., Bowman S., Baskcock K., Bankier A., Brown D.,
 RA Chillingworth T., Connor R., Devlin K., Gentles S., Hamlin N.,
 RA Harris D.E., Horsnell T., Hunt S., Jagels K., Jones M., Lye G.,
 RA Moutle S., Odell C., Pearson D., Rajandream M.A., Rice P., Rowley N.,
 RA Skelton J., Smith V., Walsh S., Whitehead S., Barrett B.G.;
 RT "The nucleotide sequence of *Saccharomyces cerevisiae* chromosome IX.";
 RL Nature 387:84-87(1997).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: SOME, TO S.POMBE SPAC23C11.01.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; 246728; CAAB6704.1; -.
 DR PIR; S49790; S49790.
 DR Germonline; 139625; -.
 DR SGD; S0001352; YII090W.
 KM Hypothetical protein; Transmembrane.
 FT TRANSMEM 14 34 POTENTIAL.
 FT TRANSMEM 41 61 POTENTIAL.
 FT TRANSMEM 75 95 POTENTIAL.
 FT TRANSMEM 150 170 POTENTIAL.
 FT TRANSMEM 185 205 POTENTIAL.
 FT TRANSMEM 219 239 POTENTIAL.
 FT TRANSMEM 295 315 POTENTIAL.
 FT TRANSMEM 374 394 POTENTIAL.
 FT TRANSMEM 414 434 POTENTIAL.
 FT TRANSMEM 462 482 POTENTIAL.
 SQ SEQUENCE 491 AA; 56332 MW; BAEC65534E04239 CRC64;

 Alignment Scores:
 Pred. No.: 312 Length: 491
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best local Similarity: 100.00% Mismatches: 0
 Query Match: 1.09% Indels: 0
 DB: 1 Gaps: 0

 US-10-029-345A-108_COPY_532_2532 (1-2001) x YII0_YEAST (1-491)
 QY 83 ACTTTTCGTCGTCACCTTCC 63
 DB 353 ThpSeraValProLeuSer 359

 RESULT 33
 FLIC_SALTY STANDARD; PRT; 494 AA.
 AC P06179; P97160; Q02871; Q56088;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Flagellin (Phase-1-I flagellin).
 GN FLIC OR FLAP OR HMG OR STM1959.
 OS *Salmonella typhimurium*.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; *Salmonella*.
 OC NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86059460; PubMed=2999134;
 RA Jøys T.M.;
 RT "The covalent structure of the phase-1 flagellar filament protein of
 RT *Salmonella typhimurium* and its comparison with other flagellins.";
 RL J. Biol. Chem. 260:15758-15761(1985).

RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93139045; PubMed=8423149;
 RA Okazaki N., Matsuo S., Saito K., Tomimaga A., Enomoto M.;
 RT "Conversion of the *Salmonella* phase 1 flagellin gene fljB to the
 RT phase 2 gene fljB on the *Escherichia coli* K-12 chromosome.";
 RL J. Bacteriol. 175:758-766(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of *Salmonella enterica* serovar *Typhimurium*
 RT LT2.";
 RL Nature 413:852-856(2001).
 RN [4]
 RP SEQUENCE OF 1-20 FROM N.A.
 RX MEDLINE=90294238; PubMed=2193164;
 RA Homma M., Derossier D.J., McNab R.M.;
 RT "Flagellar hook and hook-associated proteins of *Salmonella*
 RT *typhimurium* and their relationship to other axial components of the
 RT flagellum.";
 RL J. Mol. Biol. 213:819-832(1990).
 RN [5]
 RP SEQUENCE OF 1-19 FROM N.A.
 RX MEDLINE=81054741; PubMed=6933466;
 RA Zeig J., Simon M.;
 RT "Analysis of the nucleotide sequence of an invertible controlling
 RT element.";
 RL Proc. Natl. Acad. Sci. U.S.A. 77:4196-4200(1980).
 RN [6]
 RP SEQUENCE OF 144-427 FROM N.A.
 RX MEDLINE=90130292; PubMed=2404944;
 RA Smith N.H., Seldner R.K.;
 RT "Sequence invariance of the antigen-coding central region of the phase
 RT 1 flagellar filament gene (fljC) among strains of *Salmonella*
 RT *typhimurium*.";
 RL J. Bacteriol. 172:603-609(1990).
 RN [7]
 RP SEQUENCE OF 475-494 FROM N.A.
 RC STRAIN=LT2 / ATCC 23564;
 RX MEDLINE=97311993; PubMed=9168604;
 RA Burnens A.P., Stanley J., Sack R., Hunziker P., Brodard I.,
 RA Nicolet J.;
 RT "The flagellin N-methylase gene fljB and an adjacent serovar-specific
 RT IS200 element in *Salmonella typhimurium*.";
 RL Microbiology 143:1539-1547(1997).
 CC -1- FUNCTION: Flagellin is the subunit protein which polymerizes to
 CC form the filaments of bacterial flagella.
 CC -1- MISCELLANEOUS: Individual *salmonella* serotypes usually alternate
 CC between the production of 2 antigenic forms of flagella, termed
 CC phase-1 and phase-2, each specified by separate structural genes,
 CC fljC AND fljB.
 CC -1- SIMILARITY: Belongs to the bacterial flagellin family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M11332; AAA27072.1; -.
 DR EMBL; D13689; BAA02846.1; -.
 DR EMBL; AE008787; AAL20871.1; -.
 DR EMBL; X51740; CAA6029.1; -.
 DR EMBL; J01801; AAA27074.1; -.
 DR EMBL; M33808; AAA27080.1; -.


```

DR EMBL; 254217; CAA09050.1; -.
DR PIR; A24262; A24262.
DR PIR; S16121; S16121.
DR PDB; 1IO1; 15-APR-03.
DR StyGene; SG10115; FLIC.
DR InterPro; IPR001029; Flagellin C.
DR InterPro; IPR001492; Flagellin N.
DR Pfam; PF00700; Flagellin C.1.
DR Pfam; PF00669; Flagellin N.1.
DR PRINTS; PRO0207; FLAGELLIN.
DR Flagellum; Complete proteome; 3D-structure.
KW INIT MET
FT INIT MET
FT CONFLICT 126 0
FT CONFLICT 126 0
FT CONFLICT 132 126
FT CONFLICT 132 126
FT CONFLICT 214 132
FT CONFLICT 260 272
FT CONFLICT 276 283
FT CONFLICT 286 290
FT CONFLICT 337 337
FT CONFLICT 345 345
FT CONFLICT 353 353
FT CONFLICT 374 376
FT CONFLICT 381 381
FT CONFLICT 386 386
FT CONFLICT 389 389
FT CONFLICT 393 393
FT CONFLICT 399 399
FT CONFLICT 402 402
FT CONFLICT 447 447
SQ SEQUENCE 494 AA; 51480 MW; 599EAD0DEA9DBEB8 CRC64;

```

```

Alignment Scores:
Pred. No.: 312
Score: 7.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 1.09%
Matches: 7
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

```

```

US-10-029-345A-108_COPY_532_2532 (1-2001) x FLIC_SALTY (1-494)

```

```

OY 1743 GTAACGCACTGCTCGTA 1723
DB 300 ValThrglyThrAlaSerVal 306

```

RESULT 34

```

RNPL_HUMAN STANDARD; PRT; 494 AA.
AC Q9H4U8; Q9H033; Q9NV00;

```

```

DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Arginyl aminopeptidase-like 1 (EC 3.4.11.-) (RNPEP-like protein).
GN RNPEP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

```

```

RP SEQUENCE FROM N.A. AND TISSUE SPECIFICITY.
MEDLINE=20472315; PubMed=11017071;

```

```

RA Horikawa Y., Oda N., Cox N.J., Li X., Orho-Melander M., Hara M.,
RA Hinojosa Y., Lindner T.H., Washima H., Schwarz P.E.H.,
RA del Bogue-Platawa L., Horikawa Y., Oda Y., Yoshitachi I., Collila S.,
RA Polonsky K.S., Wei S., Concannon P., Iwasaki N., Schlitz J.,
RA Baler L.J., Bogardus C., Groop L., Boerwinkle E., Hais C.L.,
RA Bell G.I.;

```

```

RT "Genetic variation in the calpain 10 gene (CAPN10) is associated with
RT type 2 diabetes.";
RL Nat. Genet. 26:163-175(2000).

```

```

RP SEQUENCE OF 53-494 FROM N.A.
TISSUE=Pancreas;

```

```

RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Dichtchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo W.F., Caevalier T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loggiano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Ketterman M., Madan A., Rodriguez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA Butterfield J.S.N., Krzywinski M.I., Skalska U., Smalley D.E.,
RA Schmeckel A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

```

```

RP SEQUENCE OF 116-494 FROM N.A.

```

```

RA Isozaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 242-494 FROM N.A.

```

```

RA Bloecher H., Boecher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;
RT Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 116-494 FROM N.A.

```

```

RA Isozaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 242-494 FROM N.A.

```

```

RA Bloecher H., Boecher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;
RT Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 116-494 FROM N.A.

```

```

RA Isozaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 242-494 FROM N.A.

```

```

RA Bloecher H., Boecher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;
RT Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 116-494 FROM N.A.

```

```

RA Isozaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 242-494 FROM N.A.

```

```

RA Bloecher H., Boecher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;
RT Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 116-494 FROM N.A.

```

```

RA Isozaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 242-494 FROM N.A.

```

```

RA Bloecher H., Boecher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;
RT Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 116-494 FROM N.A.

```

```

RA Isozaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 242-494 FROM N.A.

```

```

RA Bloecher H., Boecher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;
RT Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 116-494 FROM N.A.

```

```

RA Isozaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 242-494 FROM N.A.

```

```

RA Bloecher H., Boecher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;
RT Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 116-494 FROM N.A.

```

```

RA Isozaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 242-494 FROM N.A.

```

```

RA Bloecher H., Boecher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;
RT Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 116-494 FROM N.A.

```

```

RA Isozaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 242-494 FROM N.A.

```

```

RA Bloecher H., Boecher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;
RT Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 116-494 FROM N.A.

```

```

RA Isozaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 242-494 FROM N.A.

```

```

RA Bloecher H., Boecher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;
RT Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 116-494 FROM N.A.

```

```

RA Isozaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 242-494 FROM N.A.

```

```

RA Bloecher H., Boecher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;
RT Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 116-494 FROM N.A.

```

```

RA Isozaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 242-494 FROM N.A.

```

```

RA Bloecher H., Boecher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;
RT Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 116-494 FROM N.A.

```

```

RA Isozaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 242-494 FROM N.A.

```

```

RA Bloecher H., Boecher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;
RT Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 116-494 FROM N.A.

```

```

RA Isozaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 242-494 FROM N.A.

```

```

RA Bloecher H., Boecher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;
RT Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 116-494 FROM N.A.

```

```

RA Isozaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 242-494 FROM N.A.

```

```

RA Bloecher H., Boecher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;
RT Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 116-494 FROM N.A.

```

```

RA Isozaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 242-494 FROM N.A.

```

```

RA Bloecher H., Boecher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;
RT Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 116-494 FROM N.A.

```

```

RA Isozaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 242-494 FROM N.A.

```

```

RA Bloecher H., Boecher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;
RT Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 116-494 FROM N.A.

```

```

RA Isozaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 242-494 FROM N.A.

```

```

RA Bloecher H., Boecher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;
RT Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 116-494 FROM N.A.

```

```

RA Isozaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 242-494 FROM N.A.

```

```

RA Bloecher H., Boecher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;
RT Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 116-494 FROM N.A.

```

```

RA Isozaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 242-494 FROM N.A.

```

```

RA Bloecher H., Boecher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;
RT Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 116-494 FROM N.A.

```

```

RA Isozaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 242-494 FROM N.A.

```

```

RA Bloecher H., Boecher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;
RT Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 116-494 FROM N.A.

```

```

RA Isozaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 242-494 FROM N.A.

```

```

RA Bloecher H., Boecher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;
RT Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 116-494 FROM N.A.

```

```

RA Isozaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 242-494 FROM N.A.

```

```

RA Bloecher H., Boecher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;
RT Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 116-494 FROM N.A.

```

```

RA Isozaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 242-494 FROM N.A.

```

```

RA Bloecher H., Boecher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;
RT Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 116-494 FROM N.A.

```

```

RA Isozaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 242-494 FROM N.A.

```

```

RA Bloecher H., Boecher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;
RT Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 116-494 FROM N.A.

```

```

RA Isozaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 242-494 FROM N.A.

```

```

RA Bloecher H., Boecher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;
RT Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 116-494 FROM N.A.

```

```

RA Isozaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.

```

```

RP SEQUENCE OF 242-494 FROM N.A.

```

```

RA Bloecher H., Boecher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;
RT Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.

```


Best Local Similarity: 100.00%
Query Match: 1.09%
DB: 1
Indels: 0
Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x RNPL_HUMAN (1-494)

QY 1445 TGGCTGTGAGGCTGCGC 1425

DB 142 TripeSerClnClyLeuAla 148

RESULT 35

ID HUTH_THEAC STANDARD; PRT; 496 AA.

AC 09H16;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Probable histidine ammonia-lyase (EC 4.3.1.3) (Histidase).

GN HUTH OR TA0242.

OS Thermoplasma acidophilum.

OC Archaea; Euryarchaeota; Thermoplasmatia; Thermoplasmales;

OC Thermoplasmataceae; Thermoplasma.

OX NCBI_TaxID=2303;

RP SEQUENCE FROM N.A.

RC STRAIN=DSM 1728;

RX MEDLINE=20479972; PubMed=11029001;

RA Ruepp A., Grael W., Santos-Martinez M.-L., Koretke K.K., Volker C.,

RA Mewes H.-W., Fishman D., Stocker S., Lupas A.N., Baumeister W.;

RT "The genome sequence of the thermophilic scavenger Thermoplasma

RT acidophilum.";

RL Nature 407:508-513(2000).

CC -1- CATALYTIC ACTIVITY: L-histidine = urocanate + NH(3).

CC -1- PATHWAY: Histidine degradation; first step.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).

CC -1- PFM: Contains an active site 4-methylidene-imidazole-5-one (MIO),

CC which is formed autocatalytically by cyclization and dehydration

CC of residues Ala-Ser-Gly (By similarity).

CC -1- SIMILARITY: Belongs to the PAL / histidase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC EMBL; AL445063; CAC11387.1; -.

DR HSSP; P21310; 1B8F.

DR HAMAP; MF_00229; -1.

DR InterPro; IPR005921; Huth.

DR InterPro; IPR008948; L-Asparaginase-like.

DR InterPro; IPR001106; Ph/His_NH3lyase.

DR Pfam; PF00221; PAL; 1.

DR TIGRfam; TIGR01225; huth.1.

DR PROSITE; PS00488; PAL_HISTIDASE; FALSE_NEG.

DR Lyase; Histidine metabolism; Complete procrome.

KT CROSSLINK 141 143 5-imidazoleone (Ala-Gly)

FT MOD_RES 142 142 (By similarity).

FT MOD_RES 142 142 DHA (2,3-DIHYDROXALANINE)

FT MOD_RES 142 142 (BY SIMILARITY).

SQ SEQUENCE 496 AA; 54148 MW; D5P0BF275477CA23 CRC64;

Alignment Scores:

Pred. No.: 312

Score: 7.00

Percent Similarity: 100.00%

Best Local Similarity: 100.00%

Query Match: 1.09%

DB: 1

US-10-029-345A-108_COPY_532_2532 (1-2001) x HUTH_THEAC (1-496)

QY 1383 ACTGGTTTCGAGAGTCTGCTC 1363
DB 56 InrGlyPneGlySerLeuLeu 62

RESULT 36

ID MATK_LIOFO STANDARD; PRT; 503 AA.

AC 098370;

DT 10-OCT-2003 (Rel. 42, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Maturase K (Intron maturase).

GN MATK.

OS Liquidambar formosana (Formosan gum).

OG Chloroplast.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Saxifragales; Hamamelidaceae; Liquidambar.

OX NCBI_TaxID=63359;

RN [1]

RP SEQUENCE FROM N.A.

RA Li J.-H., Bogle A.L., Klein A.S.;

RA "Interspecific relationships and molecular divergence of Hamamelis and

RT Liquidambar (Hamamelidaceae).";

RT Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA Shi S., Huang Y., Zhang Q., Parks C.R., Wen J.;

RA "Phylogenetic inference in Altingioideae (Hamamelidaceae) based on

RT matk and its sequences.";

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: Probably assists in splicing chloroplast group II

CC introns (By similarity).

CC -1- SIMILARITY: Belongs to the intron maturase family 2. Matk

CC subfamily.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC EMBL; AF015650; AAD10955.1; -.

DR EMBL; AF133221; AAG23349.1; -.

DR InterPro; IPR000442; Intron_maturase2.

DR InterPro; IPR002866; Matk_N.

DR Pfam; PF01348; Intron_maturase2; 1.

DR Pfam; PF01824; Matk_N; 1.

KW mRNA processing; Chloroplast.

SQ SEQUENCE 503 AA; 59125 MW; C7C6A36D9B85CF35 CRC64;

Alignment Scores:

Pred. No.: 311

Score: 7.00

Percent Similarity: 100.00%

Best Local Similarity: 100.00%

Query Match: 1.05%

DB: 1

US-10-029-345A-108_COPY_532_2532 (1-2001) x MATK_LIOFO (1-503)

QY 1574 CCAGCAGCAGCAGCAGCA 1594

DB 475 ProAlaSerSerTherSarg 481

RESULT 37

ID MATK_LIOOR STANDARD; PRT; 503 AA.

AC 08WJMS; O98371;

DT 28-FEB-2003 (Rel. 41, Created)

```

DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Maturase K (Intron maturase)
GN MATK.
OS Liquidambar orientalis (Oriental sweet gum).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Saxifragales; Hamamelidaceae; Liquidambar.
OX NCBI_TaxID=63360;
RN [1]
RP SEQUENCE FROM N.A.
RA Li J.-H., Bogle A.L., Klein A.S.;
RT "Interspecific relationships and molecular divergence of Hamamelis and
RT Liquidambar (Hamamelidaceae).";
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Shi S., Huang Y., Zhang Q., Boufford D.E., Parks C.R., Wen J.;
RT "Phylogeny, evolution and biogeography of Altingioideae
RT (Hamamelidaceae) 1. Phylogenetic analyses of cpDNA matK, IGS, and
RT rDNA ITS sequences.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Probably assists in splicing chloroplast group II
CC introns (By similarity).
CC -1- SIMILARITY: Belongs to the intron maturase family 2. MatK
CC subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF015651; AAD10956.1; -
DR EMBL; AF304519; AAL56809.1; -
DR InterPro; IPR000442; Intron_maturase2.
DR InterPro; IPR002866; MatK_N.
DR Pfam; PF03348; Intron_maturase2; 1.
DR Pfam; PF01824; MatK_N; 1.
KM mRNA processing; Chloroplast.
FT CONFLICT 11 11 Y -> D (IN REF. 1).
FT CONFLICT 93 93 I -> M (IN REF. 1).
FT CONFLICT 271 271 T -> P (IN REF. 1).
FT CONFLICT 307 307 N -> H (IN REF. 1).
FT CONFLICT 333 333 I -> L (IN REF. 1).
FT CONFLICT 489 489 W -> L (IN REF. 1).
SQ SEQUENCE 503 AA; 59073 MW; 06B69C0CA015807B CRC64;

Alignment Scores:
Pred. No.: 311
Score: 7.00 Length: 503
Percent Similarity: 100.00% Matches: 7
Best Local Similarity: 100.00% Conservative: 0
Query Match: 1.05% Mismatches: 0
DB: Indels: 0
Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x MATK_LIOOR (1-503)
QY 1574 CCAGCGACGACGACCTCAGCA 1594
DB 475 ProAlaSerSerThrSerArg 481

RESULT 38
MATK_LIOOR
ID MATK_LIOOR STANDARD; PRT; 503 AA.
AC Q9G0R3; Q98372;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Maturase K (Intron maturase).

```

```

GN MATK.
OS Liquidambar styraciflua (Amberboom) (Sweetgum tree).
OC Chloroplast
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Saxifragales; Hamamelidaceae; Liquidambar.
OX NCBI_TaxID=4400;
RN [1]
RP SEQUENCE FROM N.A.
RA Li J.-H., Bogle A.L., Klein A.S.;
RT "Interspecific relationships and molecular divergence of Hamamelis and
RT Liquidambar (Hamamelidaceae).";
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=121, and 571;
RC STRAIN=121, and 571;
RA Shi S., Huang Y., Zhang Q., Parks C.R., Wen J.;
RT "Phylogenetic inference in Altingioideae (Hamamelidaceae) based on
RT matK and its sequences.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Probably assists in splicing chloroplast group II
CC introns (By similarity).
CC -1- SIMILARITY: Belongs to the intron maturase family 2. MatK
CC subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF015652; AAD10957.1; -
DR EMBL; AF133218; AAG23346.1; -
DR EMBL; AF133219; AAG23347.1; -
DR InterPro; IPR000442; Intron_maturase2.
DR InterPro; IPR002866; MatK_N.
DR Pfam; PF03348; Intron_maturase2; 1.
DR Pfam; PF01824; MatK_N; 1.
KM mRNA processing; Chloroplast.
FT CONFLICT 14 14 R -> G (IN REF. 1).
FT CONFLICT 79 79 S -> Q (IN REF. 1).
FT CONFLICT 82 82 S -> F (IN REF. 1).
FT CONFLICT 227 227 T -> N (IN REF. 1).
FT CONFLICT 409 409 K -> R (IN REF. 1).
SQ SEQUENCE 503 AA; 59099 MW; F532A586A3EC605 CRC64;

Alignment Scores:
Pred. No.: 311
Score: 7.00 Length: 503
Percent Similarity: 100.00% Matches: 7
Best Local Similarity: 100.00% Conservative: 0
Query Match: 1.05% Mismatches: 0
DB: Indels: 0
Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x MATK_LIOST (1-503)
QY 1574 CCAGCGACGACGACCTCAGCA 1594
DB 475 ProAlaSerSerThrSerArg 481

RESULT 39
MS14_ARATH
ID MS14_ARATH STANDARD; PRT; 507 AA.
AC Q22607; Q42323; Q93VE7; Q93SL1;
DT 15-UTL-1998 (Rel. 36, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE WD-40 repeat protein MS14.
GN MS14 OR AT2G19520 OR F3P11.12.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC Euroside II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Maeson T.M., Bowman C.L., Barnstead M.E., Feldlyum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,
RA Moffet K.S., Cronin L.A., Shen M., Pal G., Van Aken S., Umayam L.,
RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA Niernan W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.,
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana."
RL Nature 402:761-768(1999).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=22954850; PubMed=14593172;
RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,
RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,
RA Karlin-Nemman G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,
RA Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,
RA Chan M.W., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,
RA Arakawa T., Barn J., Banno F., Bowser L., Brooke S.Y., Carrinai P.,
RA Chao Q., Choy N., Enji A., Goldsmith A.D., Gurjal M., Hansen N.F.,
RA Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,
RA Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,
RA Kamita A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,
RA Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,
RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.,
RT "Empirical analysis of transcriptional activity in the Arabidopsis
RT genome."
RL Science 302:842-846(2003).
RN (3)
RP SEQUENCE OF 4-507 FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=99087420; PubMed=9872415;
RA Kenzior A.L., Folk W.R.,
RT "AtMS1a and Rbap48 WD-40 repeat proteins bind metal ions."
RL FEMS Lett. 440:425-429(1998).
RN (4)
RP SEQUENCE OF 1-142 AND 452-507 FROM N.A.
RC STRAIN=cv. Columbia; TISSUE=green silique;
RA Raynal M., Grellet F., Landie M., Meyer Y., Cooke R., Deleeny M.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- SIMILARITY: Belongs to the WD-repeat RBAp6/RBAp48/MS1 family.
CC -1- SIMILARITY: Contains 6 WD repeats.
CC -1- CAUTION: Ref.4 (CAA85542) sequence differs from that shown due to
CC a frameshift in position 136.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
CC or send an email to license@ebi.ac.uk).
CC
CC EMBL, AC005917; AAD10151.2; ALT_INIT.
CC
CC EMBL, AY059799; AAL24281.1; -.
CC EMBL, AY057655; AAL15286.1; -.
CC EMBL, AY081447; AAM10009.1; -.
CC EMBL, AF028711; AAD03340.1; -.
CC EMBL, Z37286; CAA85542.1; ALT_FRAME.
CC EMBL, Z37287; CAA85543.1; -.
CC PIR, G84577; G84577.
CC InterPro: IPR001680; WD40.
CC Pfam, PF00400; WD40; 5.

DR PRINTS; PR00320; GPROTEINRPT.
DR Prodom; PD000018; WD40; 1.
DR SMART; SM00320; WD40; 5.
DR PROSITE; PS00678; WD_REPEATS_1; FALSE_NEG.
DR PROSITE; PS50082; WD_REPEATS_2; 3.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat; Nuclear protein.
FT REPEAT 162 193
FT REPEAT 217 248
FT REPEAT 290 321
FT REPEAT 335 366
FT REPEAT 384 415
FT REPEAT 439 477
FT DOMAIN 468 471
FT CONFLICT 89 89
FT CONFLICT 126 126
FT CONFLICT 202 202
FT CONFLICT 270 270
FT CONFLICT 463 463
FT CONFLICT 489 489
FT CONFLICT 494 494
SQ SEQUENCE 507 AA; 55759 MW; C37F8000FB83397 CRC64;
Alignment Scores:
Pred. No.: 311
Score: 7.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 1.05%
DB: 1
Gaps: 0
US-10-029-345a-108_COPY_532_2532 (1-2001) x MS1A_ARATH (1-507)
QY 910 ACTGAGCATCAGCGCCAAAG 930
Db 20 ThrGlyAlaSerGlyProlys 26
RESULT 40
PURI_CHICK STANDARD; PRT; 510 AA.
AC P28173;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Amino-phosphoribosyltransferase precursor (EC 2.4.2.14) (Glutamine
DE phosphoribosylpyrophosphate amidotransferase) (ATASE) (GPAT).
GN GPAT OR GPAT.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=91065928; PubMed=8336716;
RA Zhou G., Dixon J.E., Zaikin H.;
RT "Cloning and expression of avian glutamine
RT phosphoribosylpyrophosphate amidotransferase. Conservation of a
RT bacterial propyl peptide sequence supports a role for posttranslational
RT processing."
RL J. Biol. Chem. 265:21152-21159(1990).
RN (2)
RP SEQUENCE OF 1-42 FROM N.A.
RX MEDLINE=9330273; PubMed=8336716;
RA Gavallas A., Dixon J.E., Brayton K.A., Zaikin H.;
RT "Coexpression of two closely linked avian genes for purine nucleotide
RT synthesis from a bidirectional promoter."
RL Mol. Cell. Biol. 13:4784-4792(1993).
RN (3)
RP ERRATUM.
RA Gavallas A., Dixon J.E., Brayton K.A., Zaikin H.;
RL Mol. Cell. Biol. 13:7977-7977(1993).
CC -1- CATALYTIC ACTIVITY: 5-phospho-beta-D-riboseylamine + diphosphate +

```

CC      L-glutamate = L-glutamine + 5-phospho-alpha-D-ribose 1-diphosphate
CC      + H(2)O.
CC      -1- COFACTOR: Binds 1 magnesium ion and 1 4Fe-4S cluster per subunit
CC      (By similarity).
CC      -1- PATHWAY: De novo purine biosynthesis, first step.
CC      -1- SUBUNIT: Homotetramer.
CC      -1- SIMILARITY: In the C-terminal section, belongs to the
CC      purine/pyrimidine phosphoribosyltransferase family.
CC      -1- SIMILARITY: Contains 1 type-2 glutamine amidotransferase domain.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; M6069; AAA62736.1; -.
CC      EMBL; L12533; AAA17895.1; -.
CC      PIR; A38337; A38337.
CC      HSSP; P00497; 1A00.
CC      MEROPS; C44.001; -.
CC      InterPro; IPR005854; Amd_phepho_trans.
CC      InterPro; IPR000583; GATase_2.
CC      InterPro; IPR002375; Pr/PY_ip_transf.
CC      InterPro; IPR000836; PRTtransferase.
CC      Pfam; PR00310; GATase_2; 1.
CC      Pfam; PR00156; Ribosyltran; 1.
CC      TIGRfam; TIGR01134; purf; 1.
CC      PROSITE; PS00103; PUR_PYR_PR_TRANSFER; 1.
CC      PROSITE; PS00443; GATASE_TYPE_II; 1.
CC      Purine biosynthesis; Transferase; Glycosyltransferase;
CC      Glutamine amidotransferase; Allosteric enzyme; Metal-binding;
CC      Magnesium; Iron-sulfur; 4Fe-4S.
CC      KEGG; K01101; PROBABLE.
CC      PROPEP; 1; 11.
CC      FT CHAIN 12 510 AMIDOPHOSPHORIBOSYLTRANSFERASE.
CC      FT ACT SITE 12 12 GATASE (BY SIMILARITY).
CC      FT METAL 280 280 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
CC      FT METAL 327 327 MAGNESIUM (BY SIMILARITY).
CC      FT METAL 389 389 MAGNESIUM (BY SIMILARITY).
CC      FT METAL 390 390 MAGNESIUM (BY SIMILARITY).
CC      FT METAL 426 426 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
CC      FT METAL 496 496 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
CC      FT METAL 499 499 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
CC      SQ SEQUENCE 510 AA; 56257 MW; F4371FB1FBC7C744 CRC64;

Alignment Scores:
Pred. No.: 311 Length: 510
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.05% Indels: 0
Gaps: 0
DB: 1

US-10-029-345A-108_COPY_532_2532 (1-2001) x PUR1_CHICK (1-510)
QY 1786 GGTCTATTCTGCGCGCGGCGG 1806
DB 299 VAlTYzSerValATgATgATg 305

RESULT 41
HUTH_VIBCH STANDARD; PRT; 511 AA.
ID HUTH_VIBCH STANDARD; PRT; 511 AA.
AC Q9KSO4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Histidine ammonia-lyase (EC 4.3.1.3) (Histidase).
GN HUTH OR VC1202.
OS Vibrrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrrio.

```

```

OX NCBI_TaxId=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / Serotype O1;
RA MEDLINE=20406833; PubMed=10952501;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Baas S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Frazer C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483 (2000).
CC -1- CATALYTIC ACTIVITY: L-histidine = urocanate + NH(3).
CC -1- PATHWAY: Histidine degradation; first step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -1- PTM: Contains an active site 4-methylidene-imidazole-5-one (MIO),
CC which is formed autocatalytically by cyclization and dehydration
CC of residues Ala-Ser-Gly (By similarity).
CC -1- SIMILARITY: Belongs to the PAL / histidase family.
CC -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL; AE004200; AAF94361.1; -.
CC      PIR; B82228; B82228.
CC      HSSP; P21310; 1B8F.
CC      TIGR; VC1202; -.
CC      DR HAMAP; MF_00229; -. 1.
CC      DR InterPro; IPR005921; Hutch.
CC      DR InterPro; IPR008948; L-Aspartase-like.
CC      DR InterPro; IPR001106; Phe/His_NH3lyase.
CC      Pfam; PR00221; PAL; 1.
CC      TIGRfam; TIGR01225; hutch; 1.
CC      PROSITE; PS00488; PAL_HISTIDASE; 1.
CC      Lyase; Histidine metabolism; Complete proteome.
CC      CROSSLINK 143 145 5-imidazolone (Ala-Gly)
CC      FT MOD_RES 144 144 DHA (2,3-DIDEHYDROALANINE)
CC      FT SEQUENCE 511 AA; 54816 MW; E872E0CB79FAF38A CRC64;

Alignment Scores:
Pred. No.: 311 Length: 511
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.05% Indels: 0
Gaps: 0
DB: 1

US-10-029-345A-108_COPY_532_2532 (1-2001) x HUTH_VIBCH (1-511)
QY 1853 AGAGCCCTTTGAAAGCAGT 1873
DB 453 ArgAlaPProLeuLYSerSer 459

RESULT 42
COXI_MYXGL STANDARD; PRT; 517 AA.
ID COXI_MYXGL STANDARD; PRT; 517 AA.
AC 021079;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Cytochrome c oxidase polypeptide I (BC 1.9.3.1).
GN COXI OR COI.
OS Myxine glutinosa (Atlantic hagfish).

```

OG Mitochondrion.
 OC Eukaryota; Metazoa; Chordata; Craniata; Hyperotreti; Myxiniiformes;
 OC Myxiniidae; Myxiniinae; Myxine.
 OX NCBI_TaxID=7769;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98210228; PubMed=9541532;
 RA Rasmussen A.S., Janke A., Arnason U.;
 RT "The mitochondrial DNA molecule of the hagfish (Myxine glutinosa) and
 vertebrate phylogeny.";
 RL J. Mol. Evol. 46:382-388(1998).
 RN (2)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21534745; PubMed=11677623;
 RA Delarbre C., Rasmussen A.S., Arnason U., Gachelin G.;
 RT "The complete mitochondrial genome of the hagfish *Myxine glutinosa*:
 unique features of the control region.";
 RL J. Mol. Evol. 53:634-641(2001).
 RN (3)
 RP SEQUENCE OF 1-8 FROM N.A.
 RX MEDLINE=97398704; PubMed=9254918;
 RA Delarbre C., Barriel V., Tillet S., Janvier P., Gachelin G.;
 RT "The main features of the craniate mitochondrial DNA between the NDI
 and the COI genes were established in the common ancestor with the
 lancelet.";
 RL Mol. Biol. Evol. 14:807-813(1997).
 CC -1- FUNCTION: Cytochrome c oxidase is the component of the respiratory
 chain that catalyzes the reduction of oxygen to water. Subunits
 1-3 form the functional core of the enzyme complex. CO I is the
 catalytic subunit of the enzyme. Electrons originating in
 cytochrome c are transferred via the copper A center of subunit 2
 and heme A of subunit 1 to the bimetallic center formed by heme A3
 and copper B.
 CC -1- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) = 4 ferricytochrome
 c + 2 H(2)O.
 CC -1- PATHWAY: Respiratory chain; terminal step.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
 inner membrane. Contains 12 potential transmembrane domains.
 CC -1- SIMILARITY: Belongs to the heme-copper respiratory oxidase family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL: Y15182; CAI75481.1; -;
 DR EMBL: AJ404477; CAC20651.1; -;
 DR EMBL: Y09527; CAI70718.1; -;
 DR PIR: T13818; T13818.
 DR HSSP: P00396; 20CC.
 DR InterPro: IPR000883; COX1.
 DR Pfam: PF00115; COX1; 1.
 DR PRINTS: PR01165; CYCOXIDASE1.
 DR PROSITE: PS00077; COX1; 1.
 KW Oxidoreductase; Heme; Copper; Mitochondrion; Transmembrane;
 KW Respiratory chain; Inner membrane.
 FT METAL 61 61 IRON (HEME A AXIAL LIGAND) (PROBABLE).
 FT METAL 240 240 COPPER B (PROBABLE).
 FT METAL 244 244 COPPER B (PROBABLE).
 FT METAL 290 290 COPPER B (PROBABLE).
 FT METAL 291 291 COPPER B (PROBABLE).
 FT METAL 376 376 IRON (HEME A3 AXIAL LIGAND) (PROBABLE).
 FT METAL 378 378 IRON (HEME A AXIAL LIGAND) (PROBABLE).
 FT CROSSLINK 240 244 1'-histidyl-3'-tyrosine (His-Tyr)
 (By similarity).
 SQ SEQUENCE 517 AA; 57947 MW; 286F869510D532AB CRC64;

Alignment Scores:
 Pred. No.:
 Score:

310 Length: 517
 7.00 Matches: 7

Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.09% Indels: 0
 DB: 1 Gaps: 0
 US-10-029-345A-108_COPY_532_2532 (1-2001) x YHG9_YEAST (1-518)
 QY 1190 TTATTGCTGCTTCGAGCCTG 1170
 Db 111 LeuLeuLeuSerSerSerLeu 117
 RESULT 43
 YHG9_YEAST STANDARD; PRT: 518 AA.
 AC P38758;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical 57.0 kDa protein in SOD2-RPL27A intergenic region.
 GN YHR009C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycace.
 OX NCBI_TaxID=4932;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RX MEDLINE=94378003; PubMed=8091229;
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
 Du Z., Favell A., Fulton L., Gattung S., Geisel C., Kirsten J.,
 Kucaba T., Hillier L.W., Jier M., Johnston L., Langston Y.,
 Latreille P., Louis R.J., Macri C., Mardis E., Meneses S., Mouser L.,
 Nhan M., Rifkin L., Riles L., St Peter H., Trevasakis E., Vaughan K.,
 Vaudin M., Vaudin D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
 RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
 VII.";
 RL Science 265:2077-2082(1994).
 CC -1- SIMILARITY: TO S.POMBE SPAC1F5.03C AND SPAC610.06.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL: U10400; AAB68938.1; -;
 DR PIR: S46784; S46784.
 DR GerMOnline: 139326; -;
 DR SGD: S0001051; YHR009C.
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 19 35 POTENTIAL.
 SQ SEQUENCE 518 AA; 57034 MW; 362C244F0F2C9959 CRC64;

Alignment Scores:

Pred. No.: 310 Length: 518
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.05% Indels: 0
 DB: 1 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x YHG9_YEAST (1-518)

QY 1190 AGCTGAGGCTTCCTCTCTC 1210

Db 179 SerSerSerValProSerLeu 185

RESULT 44
 YAV8_SCHPO STANDARD; PRT: 521 AA.
 ID_YAV8_SCHPO

AC Q10177;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein C27P1.08 in chromosome I.
 GN SPAC27P1.08.
 OS Schizosaccharomyces pombe (Pislon yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 ON NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Vancourt G., Aert R., Robben J., Grymposprez B.,
 RA Wellens J., Vancourt G., Aert R., Robben J., Grymposprez B.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moercl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Leirach H., Wambutt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Pohl T.M.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaire V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Galliard C., Tallada V., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovski G.V., Uesery D., Bartell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe";
 RL Nature 415:871-880(2002).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: Belongs to the NRAMP family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z69368; CA93297.1; -.
 DR F.R.; T38466; T38466.
 DR GeneDB: Spombe; SPAC27P1.08; -.
 DR InterPro: IPR001046; Nrmamp.
 DR Pfam: PF01566; Nrmamp; 1.
 DR PRINTS: PR00447; NATRESASSCMP.
 DR PRODOM: PD001861; Nrmamp; 1.
 DR TIGRfams: TIGR01197; nrmamp; 1.
 KW Hypothetical protein; Transmembrane; Transport.
 FT TRANSMEM 71 91 POTENTIAL.
 FT TRANSMEM 104 124 POTENTIAL.
 FT TRANSMEM 152 172 POTENTIAL.
 FT TRANSMEM 179 199 POTENTIAL.
 FT TRANSMEM 210 230 POTENTIAL.
 FT TRANSMEM 233 253 POTENTIAL.
 FT TRANSMEM 260 280 POTENTIAL.
 FT TRANSMEM 325 345 POTENTIAL.
 FT TRANSMEM 373 393 POTENTIAL.
 FT TRANSMEM 417 437 POTENTIAL.
 FT TRANSMEM 440 460 POTENTIAL.
 FT TRANSMEM 495 515 POTENTIAL.

SQ SEQUENCE 521 AA; 57641 MW; F14271E85FD7805A CRC64;
 Alignment Scores:
 Pred. No.: 310 Length: 521
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.09% Indels: 0
 DB: 1 Gaps: 0
 US-10-029-345a-108_COPY_532_2532 (1-2001) x YAV8_SCHPO (1-521)
 QY 903 CTTAATCTTCTCTCATAGTC 883
 DB 507 LeuAmlLeuLeuLeuLeuLeu 513
 RESULT 45
 ID HEMK_RICCN STANDARD; PRT; 524 AA.
 AC 092613;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Bifunctional methyltransferase [includes: Hemk protein homolog
 DE (EC 2.1.1.-) (M.RochemKP); tRNA (guanine-N(7)-)-methyltransferase
 DE (EC 2.1.1.33) (tRNA(m7G46)-methyltransferase)].
 GN HEMK OR RCI314.
 OS Rickettsia conorii.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsiae; Rickettsia.
 ON NCBI_TaxID=781;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Malish 7;
 RX MEDLINE=21442074; PubMed=11557893;
 RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barde V.,
 RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
 RA Raoult D.;
 RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii";
 RL Science 293:2093-2098(2001).
 CC -1- FUNCTION: Catalyzes the formation of N(7)-methylguanine at
 CC position 46 (m7G46) in tRNA (by similarity).
 CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-
 CC homocysteine + tRNA containing N(7)-methylguanine.
 CC -1- SIMILARITY: In the N-terminal section, belongs to the hemk family.
 CC -1- SIMILARITY: In the C-terminal section, belongs to the
 CC methyltransferase superfamily. Trmb family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: A808678; AAL03852.1; -.
 DR HAMAP: MF_01057; fused; 1.
 DR InterPro: IPR004395; Cons_hypoth91.
 DR InterPro: IPR004556; Hemk.
 DR InterPro: IPR003358; Methyltransf_4.
 DR InterPro: IPR002052; N6_Mtase.
 DR Pfam: PF02390; SAM_bind.
 DR TIGRfams: TIGR00536; hemk_fam; 1.
 DR TIGRfams: TIGR00091; TIGR00091; 1.
 DR PROSITE: PS00092; N6_MTASE; UNKNOWN; 1.
 KW Transferase; Methyltransferase; tRNA processing; Complete proteome.
 FT DOMAIN 1 306 HEMK.
 FT DOMAIN 307 524 tRNA (GUANINE-N(7)-)-METHYLTRANSFERASE.
 SQ SEQUENCE 524 AA; 60079 MW; 3B574D6CF7DFDC28 CRC64;
 Alignment Scores:

Pred. No.: 310 Length: 524
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.05% Indels: 0
 DB: 1 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x HEMK_RICCN (1-524)

QY 538 GAGCTGATACAGCAGATGCG 558
 DB 473 GtUeu1leg1ng1n1a5ng1 479

RESULT 46
 CAP_YEAST STANDARD; PRT; 526 AA.
 ID CAP_YEAST STAN 526 AA.
 AC P17555;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Adenyl cyclase-associated protein (CAP)
 GN SRV2 OR CAP1 OR YNL138W OR N1210 OR N1838.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 NC NCBI_TaxID=4932;
 RX MEDLINE=2184942;
 RA Field J., Vojtek A., Ballester R., Bolger G., Colicelli J.,
 RA Ferguson K., Gerst J., Katsaka T., Michaeli T., Powers S., Riggs M.,
 RA Rodgers L., Wieland I., Wheland B., Wiegler M.,
 RT "Cloning and characterization of CAP, the S. cerevisiae gene encoding
 RT the 70 kd adenyl cyclase-associated protein.";
 RT Cell 61:319-327 (1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=90235282; PubMed=2158860;
 RA Fedor-Chalken M., Deechenes R.U., Broach J.R.;
 RT "SRV2, a gene required for RAS activation of adenylate cyclase in
 RT yeast.";
 RT Cell 61:329-340 (1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=96109932; PubMed=8619318;
 RA Mallet L., Buserreau F., Jacquet M.;
 RT "A 43.5 kb segment of yeast chromosome XIV, which contains MPA2,
 RT MEP2, CAP/SRV2, NNM9, FKBI/PPRI/RBP1, MOM2 and CPT1, predicts an
 RT adenosine deaminase gene and 14 new open reading frames.";
 RT Yeast 11:1195-1209 (1995).
 RL -1- FUNCTION: The N-terminal domain binds to adenyl cyclase, thereby
 CC enabling adenyl cyclase to be activated by upstream regulatory
 CC signals, such as Ras. The C-terminal domain is required for normal
 CC cellular morphology and growth control.
 CC -1- SUBCELLULAR LOCATION: Cell membrane.
 CC -1- SIMILARITY: Belongs to the CAP family.
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z46844; CAA86887.1; -
 CC EMBL: M58284; AAA63569.1; -
 CC EMBL: M32663; AAA35094.1; -
 CC EMBL: Z71414; CAA86020.1; -
 CC PIR: A34896; A34896.
 CC PDB: 1K4Z; 13-WAR-02.

DR PDB; 1KQ5; 16-JAN-02.
 DR Germonline; 143144; -
 DR SGD; S0005082; SRV2
 DR GO; GO:0007265; P:RAS protein signal transduction; IMP.
 DR InterPro; IPR001837; CAP.
 DR InterPro; IPR006599; CAP.
 DR Pfam; PF01213; CAP; 1.
 DR SMART; SM00673; CAP; 2.
 DR PROSITE; PS01089; CAP_1; 1.
 DR PROSITE; PS01089; CAP_2; 1.
 DR Membrane; 3d-structure.
 FT DOMAIN 262 300
 FT DOMAIN 277 282
 SQ SEQUENCE 526 AA; 57521 MW; 0EB4D41205E2D464 CRC64;
 POLY-PRO.

Alignment Scores:
 Pred. No.: 310 Length: 526
 Score: 7.00 Matches: 7
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.09% Indels: 0
 DB: 1 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x CAP_YEAST (1-526)

QY 252 ACTGCATCATGTCACCTT 232
 DB 485 ThrAla1leAsnValAsnLeu 491

RESULT 47
 PBP_SCHPO STANDARD; PRT; 532 AA.
 ID PBP_SCHPO STAN 532 AA.
 AC O60109; Q9C427;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Alkaline phosphatase (EC 3.1.3.1).
 GN SPBC14F5.13C.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 NC NCBI_TaxID=4896;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Guilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouras J., Peat N., Hayles J., Baker S., Baeham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Ocell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabbittowitch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares K., Sharp S.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
 RA Weljens I., Vanterrele E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel K., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Leinach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Goffeau A., Cadieu E., Dreno S., Gloux S., Lelaure V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,

RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revelante J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.,
RT "The genome sequence of Schizosaccharomyces pombe";
RL Nature 415:871-880(2002).
CC -1- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O = an
CC alcohol + phosphate.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein (potential).
CC -1- SIMILARITY: Belongs to the alkaline phosphatase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF316541; AA07179.1; -;
DR EMBL: AL023780; CA19331.1; -;
DR PIR: T39459; T39459.
DR HSSP: P00634; 1A1A.
DR GeneDB: Spombe, SPB014F5.13c; -;
DR InterPro: IPR001952; ALK_phosphatase.
DR Pfam: PF00245; alk_phosphatase; 1.
DR PRINTS: PR00113; ALKPHPTASE.
DR ProDom: PD001868; Alk_phosphatase; 1.
DR SMART: SM00098; alkPc; 1.
DR PROSITE: PS00123; ALKALINE PHOSPHATASE; 1.
KW Hydrolyase; Zinc; Magnesium; Transmembrane; Signal-anchor.
FT TRANSMEM 31 48
FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT ACT_SITE 114 114
FT CONFLICT 513 514 PS -> HC (IN REF. 1).
SQ SEQUENCE 532 AA; 58666 MW; 57A84A66926D545C CRC64;

Alignment Scores:
Pred. No.: 309 Length: 532
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.05% Indels: 0
DB: 1 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x PPB_SCHPO (1-532)
QY 1547 ACACAGCTTCCTTTTCGCC 1567
DB 168 ThrProLaserPheserAla 174

RESULT 48
TCE1_AVE5A STANDARD; PRT; 535 AA.
ID TCE1_AVE5A STANDARD; PRT; 535 AA.
AC P40412;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE T-complex protein 1, epsilon subunit (TCP-1-epsilon) (CCT-epsilon)
DE (TCP-K19).
OS Avena sativa (Oat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Avenae; Avena.
OX NCBI_TaxID=4498;
OX (1)
RN SEQUENCE FROM N.A.
RP STRAIN=cv. Dewi; TISSUE=coleoptile, and Mesocotyl;
RX MEDLINE=94085629; PubMed=7903257;
RA Ehmman B., Krenz M., Mummert E., Schaefer E.;
RT "Two tcp-1-related but highly divergent gene families exist in oat
RT encoding proteins of assumed chaperone function.";
RL FEBS Lett. 336:313-316(1993).
CC

CC -1- FUNCTION: Molecular chaperone; assist the folding of proteins upon
CC ATP hydrolysis. Known to play a role, in vitro, in the folding of
CC actin and tubulin.
CC -1- SUBUNIT: Hetero-oligomeric complex of about 850 to 900 kDa that
CC forms two stacked rings, 12 to 16 nm in diameter.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the TCP-1-chaperonin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X75777; CAAS3396.1; -;
DR PIR: S40461; S40461.
DR HSSP: P48425; 1A6D.
DR InterPro: IPR002194; Chaperonin TCP-1.
DR InterPro: IPR001844; Chaperonin Cpn60.
DR InterPro: IPR002423; Cpn60/TCP-1.
DR InterPro: IPR008950; GroEL-ATPase.
DR Pfam: PF00118; Cpn60_TCP1; 1.
DR PRINTS: PR00298; CHAPERONIN60.
DR PROSITE: PS00750; TCP1_1; 1.
DR PROSITE: PS00751; TCP1_2; 1.
DR PROSITE: PS00995; TCP1_3; 1.
KW Chaperone; ATP-binding; Multigene family.
SQ SEQUENCE 535 AA; 58899 MW; 6A8F847CA891BC32 CRC64;

Alignment Scores:
Pred. No.: 309 Length: 535
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.05% Indels: 0
DB: 1 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x TCE1_AVE5A (1-535)
QY 226 AACATAGCTTGACATTGAT 246
DB 259 LysHisLysValAspIleAsp 265

RESULT 49
TCE2_AVE5A STANDARD; PRT; 535 AA.
ID TCE2_AVE5A STANDARD; PRT; 535 AA.
AC P54411;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE T-complex protein 1, epsilon subunit (TCP-1-epsilon) (CCT-epsilon)
DE (TCP-K36).
OS Avena sativa (Oat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Avenae; Avena.
OX NCBI_TaxID=4498;
OX (1)
RN SEQUENCE FROM N.A.
RP STRAIN=cv. Dewi; TISSUE=coleoptile, and Mesocotyl;
RX MEDLINE=94085629; PubMed=7903257;
RA Ehmman B., Krenz M., Mummert E., Schaefer E.;
RT "Two tcp-1-related but highly divergent gene families exist in oat
RT encoding proteins of assumed chaperone function.";
RL FEBS Lett. 336:313-316(1993).
CC -1- FUNCTION: Molecular chaperone; assist the folding of proteins upon
CC ATP hydrolysis. Known to play a role, in vitro, in the folding of
CC actin and tubulin.
CC -1- SUBUNIT: Hetero-oligomeric complex of about 850 to 900 kDa that
CC forms two stacked rings, 12 to 16 nm in diameter.
CC


```

CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the TCP-1 chaperonin family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@1sb-sib.ch).
-----
CC EMBL; X75778; CAA5397.1; -.
CC PIR; S40462; S40462.
CC HSSP; P48424; 1A6D.
CC InterPro; IPR002194; Chaperonin TCP-1.
CC InterPro; IPR001844; Chaperonin Cpn60.
CC InterPro; IPR002423; Cpn60/TCP-1.
CC InterPro; IPR008950; GroEL-ATPase.
CC Pfam; PF00118; Cpn60 TCP1; 1.
CC PRINTS; PR00298; CHAPERONIN60.
CC PRINTS; PR00304; TCOMPLEXTCP1.
CC PROSITE; PS00750; TCP1_1; 1.
CC PROSITE; PS00751; TCP1_2; 1.
CC PROSITE; PS00995; TCP1_3; 1.
CC Chaperone; ATP-binding; Multigene family.
CC Chaperone; ATP-binding; Multigene family.
SQ SEQUENCE 535 AA; 59034 MW; 02761226F5F17B81 CRC64;

Alignment Scores:
Pred. No.: 309 Length: 535
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.05% Indels: 0
DB: 1 Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x TCE2_AVESa (1-535)

QY 226 AACATAGAGTTGACATTGAT 246
Db 259 LysHlelyValAspIleAsp 265

RESULT 50
TCPE_ARATH STANDARD; PRT; 535 AA.
AC 004450;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE T-complex protein 1, epsilon subunit (TCP-1-epsilon) (CCT-epsilon).
GN ATIG24510 OR P2139.12.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RC MEDLINE=21016719; PUBMED=1130712;
RA Theologis A., Ecker J.R., Palm C.J., Bowmen C.L., Brooke S.Y.,
RA White O., Alonso J., Altati H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Creasy T.H., Dewar K.,
RA Dunn P., Egu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gull J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hulzar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Mailli R., Marziani A.,
RA Miltchev J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salberg S.L., Schwartz J.R., Shin P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Tortumi M.J., Town C.D.,

```

```

RA Uterback T., Van Aken S., Vayberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Frazer C.M., Venter J.C., Davis R.W.;
RA "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RA thaliana."
RL Nature 408:816-820 (2000).
CC -1- FUNCTION: Molecular chaperone, assist the folding of proteins upon
CC ATP hydrolysis. Known to play a role, in vitro, in the folding of
CC actin and tubulin.
CC -1- SUBUNIT: Hetero-oligomeric complex of about 850 to 900 kDa that
CC forms two stacked rings, 12 to 16 nm in diameter.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
CC -1- SIMILARITY: Belongs to the TCP-1 chaperonin family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
CC or send an email to license@1sb-sib.ch).
-----
CC EMBL; AC000103; AAF97977.1; -.
CC HSSP; P48424; 1A6D.
CC InterPro; IPR002194; Chaperonin TCP-1.
CC InterPro; IPR001844; Chaperonin Cpn60.
CC InterPro; IPR002423; Cpn60/TCP-1.
CC InterPro; IPR008950; GroEL-ATPase.
CC Pfam; PF00118; Cpn60 TCP1; 1.
CC PRINTS; PR00298; CHAPERONIN60.
CC PRINTS; PR00304; TCOMPLEXTCP1.
CC PROSITE; PS00750; TCP1_1; 1.
CC PROSITE; PS00751; TCP1_2; 1.
CC PROSITE; PS00995; TCP1_3; 1.
CC Chaperone; ATP-binding; Multigene family.
SQ SEQUENCE 535 AA; 59384 MW; 1CB56343A3AF1DC3 CRC64;

Alignment Scores:
Pred. No.: 309 Length: 535
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.05% Indels: 0
DB: 1 Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x TCPE_ARATH (1-535)

QY 226 AACATAGAGTTGACATTGAT 246
Db 259 LysHlelyValAspIleAsp 265

Search completed: June 21, 2004, 13:10:22
Job time : 99 secs

```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 21, 2004, 12:59:47 ; Search time 102.5 Seconds

(without alignments)
12319.062 Million cell updates/sec

Title: US-10-029-345a-108_COPY_532_2532

Perfect score: 667
Sequence: 1 gttcgtcatggccatgatgac.....tggaaatcattgagctctcc 2001

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Word size: 1

Total number of hits satisfying chosen parameters: 344724

Minimum DB seq length: 473

Maximum DB seq length: 200000000

Post-processing: listing first 75 summaries

Command line parameters:

-MODE=frame+ n2p, model -DEV=xlp
-Q=/cgn2_1/USPTO.spool_p/US10029345/runat_21062004_124615_5691/app_query.fasta_1.2183
-DB=SPTRMBL_25 -QFMT=fastan -SUFFIX=oligo.rspt -MINMATCH=0.1 -LOOFC=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=75 -DOCALLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=50 -MODE=LOCAL
-OUTFMT=pro -NOR=ext -HEAPSIZ=500 -MINLEN=473 -MAXLEN=200000000
-USER=US10029345_QCGN_1_1_112_@runat_21062004_124615_5691 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

SPTRMBL_25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_verticbrate:*
14: sp_unclassified:*
15: sp_rvivirus:*
16: sp_bacteriaph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	DB	ID	Description
1	472	70.8	665	4	Q96N49		Q96N49 homo sapien

2	400	60.0	662	4	Q8IVT8	Q8IVT8 homo sapien
3	371	55.6	616	4	Q8N5T1	Q8N5T1 homo sapien
4	67	10.0	660	11	Q920R2	Q920R2 mus musculu
5	64	9.6	677	11	Q93MG6	Q93MG6 mus musculu
6	48	7.2	622	11	Q93MG5	Q93MG5 mus musculu
7	20	3.0	625	4	Q86S88	Q86S88 homo sapien
8	20	3.0	665	11	Q7RS29	Q7RS29 mus musculu
9	13	1.9	476	5	Q9VHV8	Q9VHV8 drosophila
10	13	1.9	476	5	Q46122	Q46122 drosophila
11	10	1.6	498	16	Q82Z00	Q82Z00 enterococcu
12	9	1.4	550	16	Q89P10	Q89P10 bradyrhizob
13	9	1.3	606	5	Q85T18	Q85T18 caenorhabdi
14	9	1.3	657	5	Q85T19	Q85T19 caenorhabdi
15	8	1.2	486	16	Q8BPN9	Q8BPN9 xanthomonas
16	8	1.2	495	4	Q9H6T2	Q9H6T2 homo sapien
17	8	1.2	495	4	Q96A99	Q96A99 homo sapien
18	8	1.2	497	16	Q88T27	Q88T27 lactobacill
19	8	1.2	513	10	Q86S58	Q86S58 oryza sativ
20	8	1.2	513	10	Q7XEV9	Q7XEV9 oryza sativ
21	8	1.2	513	16	Q92M58	Q92M58 rhizobium m
22	8	1.2	515	10	Q9FX19	Q9FX19 arabidopsis
23	8	1.2	525	10	Q81CP6	Q81CP6 arabidopsis
24	8	1.2	525	10	Q9LR07	Q9LR07 arabidopsis
25	8	1.2	526	12	Q65756	Q65756 bluetongue
26	8	1.2	534	12	Q9YQX9	Q9YQX9 ranid herpe
27	8	1.2	544	16	Q82GP4	Q82GP4 streptomyce
28	8	1.2	545	16	Q92ZH5	Q92ZH5 rhizobium m
29	8	1.2	555	12	Q55887	Q55887 human para
30	8	1.2	555	12	Q55888	Q55888 recombinant
31	8	1.2	559	16	Q8ABA4	Q8ABA4 bacteroides
32	8	1.2	569	16	Q7WG72	Q7WG72 bordetella
33	8	1.2	569	16	Q7W4P6	Q7W4P6 bordetella
34	8	1.2	573	16	Q7UQD7	Q7UQD7 rhodospirell
35	8	1.2	605	3	Q9PEM8	Q9PEM8 aspergillus
36	8	1.2	605	4	Q7Z6C5	Q7Z6C5 homo sapien
37	8	1.2	612	16	Q8D394	Q8D394 wigglewort
38	8	1.2	612	10	Q9ZUK4	Q9ZUK4 arabidopsis
39	8	1.2	677	4	Q8N4M1	Q8N4M1 homo sapien
40	8	1.2	682	10	Q85201	Q85201 arabidopsis
41	8	1.2	694	16	Q82K16	Q82K16 salmoneila
42	8	1.2	707	17	Q8U1R9	Q8U1R9 pyrococcus
43	8	1.2	711	12	P87583	P87583 chimpunk pa
44	8	1.2	727	16	Q8Y2P4	Q8Y2P4 anabaena sp
45	8	1.2	752	8	Q98505	Q98505 guillardi
46	8	1.2	757	2	Q9K1B0	Q9K1B0 porphyromon
47	8	1.2	759	12	Q91C22	Q91C22 ct virus. o
48	8	1.2	773	10	Q8L3X5	Q8L3X5 oryza sativ
49	8	1.2	789	16	Q8UGX6	Q8UGX6 agrobacteri
50	8	1.2	791	16	Q92RA3	Q92RA3 rhizobium m
51	8	1.2	849	16	Q8P8D5	Q8P8D5 xanthomonas
52	8	1.2	858	11	Q8K058	Q8K058 mus musculu
53	8	1.2	869	2	Q9AJ94	Q9AJ94 legionella
54	8	1.2	892	11	Q80T06	Q80T06 mus musculu
55	8	1.2	935	5	Q85WR0	Q85WR0 drosophila
56	8	1.2	1001	11	Q8BNG7	Q8BNG7 mus musculu
57	8	1.2	1037	16	Q7U036	Q7U036 rhodospirell
58	8	1.2	1042	16	Q820R6	Q820R6 nitrosomona
59	8	1.2	1065	5	Q815U3	Q815U3 plasmodium
60	8	1.2	1073	5	Q9W1T8	Q9W1T8 drosophila
61	8	1.2	1153	5	Q93598	Q93598 caenorhabdi
62	8	1.2	1194	10	Q9PKF1	Q9PKF1 arabidopsis
63	8	1.2	1222	16	Q87XX8	Q87XX8 pseudomonas
64	8	1.2	1208	5	Q01924	Q01924 caenorhabdi
65	8	1.2	1717	11	Q7TSA9	Q7TSA9 mus musculu
66	8	1.2	1762	10	Q94DC2	Q94DC2 oryza sativ
67	8	1.2	1885	17	Q8TOD1	Q8TOD1 mechanosarc
68	8	1.2	1900	12	Q9W3P1	Q9W3P1 sindbis-lik
69	8	1.2	1901	12	Q88425	Q88425 sindbis-lik
70	8	1.2	2487	12	Q9YJX8	Q9YJX8 sindbis vir
71	8	1.2	2500	12	Q88431	Q88431 sindbis-lik
72	8	1.2	2513	12	Q87644	Q87644 sindbis vir
73	8	1.2	2517	12	Q9YJX6	Q9YJX6 sindbis-lik
74	8	1.2	2761	5	Q18447	Q18447 caenorhabdi

Db	614	HsGluGluSerProHehGluGlnPheLyArgArgSerGlyGlnMetGluPheGly	633
Qy	1906	GAGAGCATCATGTTCAGAGAACAGGTCAACGGGAAAGACTGGGAAAGTGGCACTGCT	196
Db	634	GuserIleMetSerGluAsnArgSerArgGluGluLeuGlyValGlySerGlnSer	653
Qy	1966	AGCTTTTGGGAGCAATGGAAATCATTTAGAGTCC	2001
Db	654	SerPheSerGlySerMetGluIleIleGluValSer	665
RESULT 2			
Q81VT8	PRELIMINARY;	PRT;	662 AA.
AC	Q81VT8;		
DT	01-MAR-2003 (TREMBLrel. 23, Created)		
DT	01-MAR-2003 (TREMBLrel. 23, Iast sequence update)		
DT	01-OCT-2003 (TREMBLrel. 25, Iast annotation update)		
DE	Similar to dual specificity phosphatase 16.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxId=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Ductenum;		
RA	Struhsberg R.;		
RL	Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; BC042101; AAH42101.1; -		
DR	GO; GO:0017017; P:MAP kinase phosphatase activity; IEA.		
DR	GO; GO:0064470; P:protein amino acid dephosphorylation; IEA.		
DR	InterPro; IPR008340; DS_phosphatase.		
DR	InterPro; IPR001763; Rhodanese-like.		
DR	InterPro; IPR000387; TYR_phosphatase.		
DR	Pfam; PF00782; DSPC; 1.		
DR	Pfam; PF00581; Rhodanese; 1.		
DR	PRINTS; PR01764; MARKPHPTASE.		
DR	SMART; SM00195; DSPC; 1.		
DR	SMART; SM00450; RHOD; 1.		
DR	PROSITE; PSS0206; RHODANSE_3; 1.		
DR	PROSITE; PSS0383; TYR_PHOSPHATASE_1; 1.		
DR	PROSITE; PSS0056; TYR_PHOSPHATASE_2; 1.		
DR	PROSITE; PSS0054; TYR_PHOSPHATASE_DUAL; 1.		
DR	SEQUENCE 662 AA; 72818 MW; 350534EF0652B98F CRC64;		
Alignment Scores:			
Pred. No.:	0	Length:	662
Percent Similarity:	400.00	Matches:	465
Best Local Similarity:	98.12%	Conservative:	0
Query Match:	59.97%	Mismatches:	9
DB:	4	Indels:	9
		Gaps:	0
DBS-10-029-345A-108_COPY_532_2532 (1-2001) x Q81VT8 (1-662)			
Qy	586	ACCTGTCCAAAGCCTGACTTTATGCCCGAGTCCATTTCCGTGCTGCTGCTGCTGAATGAC	645
Db	194	ThiCybProGlyProAspPheIleProGluSerHisPheLeuAlaGlyAlaProValAsnAsp	213
Qy	646	AGCTTTTGGAGAAATTTTGGCCGTGGTGGACAAATCAGTAGATTTCATTGGAAGAACA	705
Db	214	SerPheCyGlnLysIleLeuProTyrLeuAspLysSerValAspPheIleGlyVala	233
Qy	706	AAAGCTCCAAATGGATGTGTTCTAGTCACTGTTTAGCTGGGATCTCCGCGCCAC	765
Db	234	LysAlaSerAsnGlyValLeuValHisCybLeuAlaGlyIleSerAspSerAlaThr	253
Qy	766	ATCGCTATGCCCTACATCATGAAGAGATGAGCATGCTTTAGATGAAGCTTACAGATT	825
Db	254	IleAlaIleAlaIleTyrIleMetLysArgMetLysSerIleuAspGluAlaTyrArgPhe	273
Qy	826	GTGAAAGAAAAAGACTTACTATATGTCCAAATTCAATTTTCTGGGCCAAGTCTGGAC	885

[illegible]

QY 1966 AGCTTTTCGGGCGAGCATGGAATCATTTAGGCTCTCC 2001
Db 651 SerPheSerGlySerMetGluIleIleGluValSer 662

RESULT 3

Q8N5T1 PRELIMINARY; PRT; 616 AA.
AC Q8N5T1;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;

RP SEQUENCE FROM N.A.
RC TISSUE=Brain, and Lung;
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC031643; AAB31643.1; -
DR GO: GO:0017017; F:MAP kinase phosphatase activity; IEA.
DR CO: CO:006470; P:protein amino acid dephosphorylation; IEA.
DR InterPro: IPR000340; DS_phosphatase.
DR InterPro: IPR008343; MAPK_phosph.
DR Pfam: PF00782; DSPC; 1.
DR InterPro: IPR000387; TYR_phosphatase.
DR PRINTS: PR01764; MAPKDPHTASE.
DR SMART: SM00195; DSPC; 1.
DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE: PS50056; TYR_PHOSPHATASE_2; 1.
DR PROSITE: PS50054; TYR_PHOSPHATASE_DUAL; 1.
KM Hypothetical protein.
FT NCN TER 1
SQ SEQUENCE 616 AA; 67636 MW; 2CB0B14482F2AD72 CRC64;

Alignment Scores:
Pred. No.: 0 Length: 616
Score: 371.00 Matches: 471
Percent Similarity: 99.79% Conservative: 0
Best Local Similarity: 99.79% Mismatches: 1
Query Match: 55.62% Indels: 0
DB: 4 Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x Q8N5T1 (1-616)

QY 586 ACCGTCCTCAAGCTGCTTATCCCGAGCTCATTTCTGCGTGGCCTGTGAATGAC 645
Db 145 ThnCysProLysProAlaPheIleProGluSerHisPheLeuArgValProValaLanasp 164
QY 646 AGCTTTGTGAGAAATTTTGGCGTGTGGACAATCAGTAGATTTCATTGAGAAAGCA 705
Db 165 SerPheCysGluValIleLeuProTrpLeuAspLysSerValAspPheIleGluVala 184
QY 706 AAAGCTCCCAATGATGTCTTACGCACTGTTGCTGCGATCTCCCGCTCCGCGCAC 765
Db 185 LysIleSerAsnGlyCysValLeuValHisCysLeuIleGlyIleSerAspSerIleAla 204
QY 766 ATGCGTATCCCTCATCATGAGAGAGATGAGACATGCTCTTATGATGAAGCTTACAGATT 825
Db 205 ILeAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAla 224
QY 826 GTGAAGAAAAAAGACCTACTATATCTTCAAACTTCATTTTCTGGGCAATCTCGAGAC 885
Db 225 ValLysGluLysArgProThrIleSerProAsnPheAsnPheLeuGlyIleLeuLeuAsp 244
QY 886 TATGAGAAGAAGATTAGAACCACTGAGCATCCAGGCGCAAGAGCAATCCAGCTG 945
Db 245 TyrGluLysLysIleLysAsnGlnThrGlyIleSerIleProLysSerLysLeuValSer 264
QY 946 CTGCACCTGGAGAACCAATGAACCTGTCTCTGCTGTCTAGAGGCTGGACAGAAAAAGC 1005

Db 265 LeuHisLeuGluLysProAlaPheIleProValProAlaValSerGluGlyGlnLysSer 284
QY 1006 GAGAGCGCCCTCAGTCCAGCCCTGTGCGACTGTCTTACTTCAGAGCAGACGACAAAGC 1065
Db 285 GluThrProLeuSerProProCysAlaAspSerAlaThrSerGluAlaIleGlyLysArg 304
QY 1066 CCCGTGCATCCCGCAGCGTCCAGAGGTGCCAGCGTCCGAGCGGCTGTGAGAGAC 1125
Db 305 ProValHisProAlaSerValProSerValProSerValGlnProSerLeuLeuAsp 324
QY 1126 AGCCGCTGTGACAGGCGCTCAGTGGCGCTGACCTGTCCGAGACAGAGCTGGAGACAGC 1185
Db 325 SerProLeuValGlnAlaLeuSerGlyLeuHisLeuSerAlaAspArgLeuGluAspSer 344
QY 1186 AATAAGCTCAAGCGTCTCTCTCTGAGATATCAATCAGTTTATTTACGCCAGCATG 1245
Db 345 AsnLysLeuLysArgSerPheSerLeuAspIleLysSerValSerTySerAlaSerMet 364
QY 1246 GCAGATCTTCAATGAGCTTCTCCTCATCAGAAAGATGCTTGAATCTCAAACTTCC 1305
Db 365 AlaAlaSerLeuHisGlyPheSerSerSerGlnAspAlaLeuGluLysTyTrpLysProSer 384
QY 1306 ACTACTCTGATGGGAGCCCAACAGCTATGCCAGTTCTCCCTGTTCAGGAACTATCGAG 1365
Db 385 ThrThrLeuAspGlyThrAsnLysLeuGlyGlnPheSerProValGlnGluLeuSerGlu 404
QY 1366 CAGACTCCCGAAGCAGCTCTGATTAAGAGAGAACCCAGCATCCCAAGAACCTGCAGACC 1425
Db 405 GlnThrProGluThrSerProAspLysGlnIleAlaSerIleProLysLysLeuGlnThr 424
QY 1426 GCCAGCGCTTCAAGACAGCAGACAGAGGATTTGATTCGATGAGTGGAGCGTGGAGACAATTAC 1485
Db 425 AlaArgProSerAspSerGlnSerIleArgLeuHisSerValAlaGlnThrSerAsnSerGly 444
QY 1486 ACCGCGCAGAGCTCCCTTTATCTTCCACTGTCATGAGTGGAGCGTGGAGACAATTAC 1545
Db 445 ThrAlaGlnArgSerLeuLeuSerProLeuHisArgSerGlySerValGluAspAsnLys 464
QY 1546 CACACACACTTCTTTTGGCGCTTCCACAGCCAGCAGACCTCAGAACTGTGCTGGC 1605
Db 465 HisThrSerPheLeuPheGlyLeuSerThrSerGlnHisLysLeuThrLysSerAlaGly 484
QY 1606 CTGGGCTTAAAGGGGTGCGACTCGGATCTTGGCCCCCGCAGACTTAACTCCCTCTG 1665
Db 485 LeuGlyLeuLysGlyLysThrPheSerAspIleLeuAlaProGlnThrSerThrProSerLeu 504
QY 1666 ACCAGCAGCTGATTTTGGCCAGAGCTCTGACACTTCTACTGCTCAGCCATCTAC 1725
Db 505 ThrSerSerTrpLysPheAlaThrGluSerSerHisPheTySerAlaSerAlaIleLys 524
QY 1726 GGAGGCAATGCGAGTACTGCTGCTTACGCGAGCGCAGCGCCACTTGGGAGACAA 1785
Db 525 GlyLysSerAlaSerTySerAlaLysSerCysSerGlnLeuProThrCysGlyAspGln 544
QY 1786 GTCTATCTGTGCGAGCGCGCAGAGCCAGATGACAGAGCTGACTCGCGGAGCTGG 1845
Db 545 ValTySerValArgArgArgGlnLysProSerAspArgAlaAspSerArgSerTrp 564
QY 1846 CATGAAGAGAGCCCTTTGAAAAGCATTTTAAAGCAGAACTGTGCCAAATGGAATTTGCA 1905
Db 565 HisGluGluSerProPheGluLysGlnPheLysArgArgSerCysGlnMetGlnPheLys 584
QY 1906 GAGAGCATCAGTTCAGAGAACAGGTCAGCGGAGAGAGTGGGGAAGAGTGGGAGTCACTCT 1965
Db 585 GluSerIleLeuSerGlnAsnArgSerArgGlnLysGlyLysValGlySerGlnSer 604
QY 1966 AGCTTTTCGGGCGAGCATGAATCATTTAGGCTCTCC 2001
Db 605 SerPheSerGlySerMetGluIleIleGluValSer 616
RESULT 4 Q920R2

ID 0920R2 PRELIMINARY; PRT; 660 AA.
AC 0920R2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE MAP kinase phosphatase-7.
GN DUSP16 OR 3830417M17RIK OR MKP-7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum;
RX MEDLINE=21486429; PubMed=11489891;
RA Maeda K., Shima H., Watanabe M., Kikuchi K.,
RT "MKP-7, a Novel Mitogen-activated Protein Kinase Phosphatase,
Functions as a Shuttle Protein".
RL J. Biol. Chem. 276:39002-39011(2001).
DR EMBL: AB052157; BAB47240.1; -.
DR HSSP: Q16828; 1MKP.
DR MGD: MG1:1917936; Dusp16.
DR GO: GO:0005737; C:cytoplasm; IDA.
DR GO: GO:0005634; C:nucleus; IDA.
DR GO: GO:0005515; F:protein binding; IPI.
DR GO: GO:000188; P:inactivation of MAPK; IDA.
DR InterPro: IPR000340; DS phosphatase.
DR InterPro: IPR008343; MAPK phosph.
DR InterPro: IPR001763; Rhodanese-like.
DR InterPro: IPR000387; TYR_phosphatase.
DR Pfam: PF00782; DSPC; 1.
DR Pfam: PF00581; Rhodanese; 1.
DR SMART: SM00195; DSPC; 1.
DR SMART: SM00450; RHOD; 1.
DR PROSITE: PS50206; RHODANSE_3; 1.
DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE: PS50056; TYR_PHOSPHATASE_2; 1.
DR PROSITE: PS50054; TYR_PHOSPHATASE_DUAL; 1.
KW Hydroxylase; Kinase.
SQ SEQUENCE 660 AA; 72695 MW; DB609FCDAD4AA309 CRC64;
Alignment Scores:
Pred. No.: 1,14e-60 Length: 660
Score: 67.00 Matches: 67
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10.04% Indels: 0
DB: 11 Gaps: 0
US-10-029-345A-108_COPY_532_2532 (1-2001) x 0920R2 (1-660)
QY 49 TTGGTGGCTCTGCTGAAAGTGAACGAAAGAGTCTGTAATGATAGCCGCCATT 108
DB 15 LeuValAlaLeuLeuGluSerGlyThrGluValLeuLeuLeuLeuSerAlaGlyProPhe 34
QY 109 GTGGAATACATATCATCTCCACATTTTGAAGCCATTATATCACTGCTCCAACTTATG 168
DB 35 ValGluTyrAsnThrSerHisIleLeuGluAlaIleAsnIleAsnGlySerIleuMet 54
QY 169 AACGGAAGTTGACGACGACGAAGTGTATTAATACAGAGCTCATCCAGCATTCAGGAAA 228
DB 55 LysArgArgLeuGlnGlnIleuAspLysValLeuIleThrGluIleuIleGlnHisSerAlaLys 74
QY 229 CATAGGTTGACATGATTGC 249
DB 75 HisIleValaLeuPheLeuAspCys 81
RESULT 5
ID 099MG6 PRELIMINARY; PRT; 677 AA.
AC 099MG6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Map kinase phosphatase-M A1 isoform.
GN DUSP16 OR 3830417M17RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALE/c;
RA Matsuguchi T., Musikacharen T., Johnson T.R., Kraft A.S.,
RA Yoshikai Y.;
RT "A Novel MKP is an Important Negative Regulator of LPS-mediated JNK
Activation in Macrophages".
RL Mol. Cell. Biol. 20:6999-7009(2001).
DR EMBL: AF345951; AAK35052.1; -.
DR HSSP: Q16828; 1MKP.
DR MGD: MG1:1917936; Dusp16.
DR GO: GO:0005737; C:cytoplasm; IDA.
DR GO: GO:0005634; C:nucleus; IDA.
DR GO: GO:0005515; F:protein binding; IPI.
DR GO: GO:000188; P:inactivation of MAPK; IDA.
DR InterPro: IPR000340; DS phosphatase.
DR InterPro: IPR008343; MAPK phosph.
DR InterPro: IPR001763; Rhodanese-like.
DR InterPro: IPR00387; TYR_phosphatase.
DR Pfam: PF00782; DSPC; 1.
DR Pfam: PF00581; Rhodanese; 1.
DR SMART: SM00195; DSPC; 1.
DR SMART: SM00450; RHOD; 1.
DR PROSITE: PS50206; RHODANSE_3; 1.
DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE: PS50056; TYR_PHOSPHATASE_2; 1.
DR PROSITE: PS50054; TYR_PHOSPHATASE_DUAL; 1.
KW Hydroxylase; Kinase.
SQ SEQUENCE 677 AA; 74550 MW; 8B6D5B7096CEC2FC CRC64;
Alignment Scores:
Pred. No.: 1,63e-57 Length: 677
Score: 64.00 Matches: 64
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.60% Indels: 0
DB: 11 Gaps: 0
US-10-029-345A-108_COPY_532_2532 (1-2001) x 099MG6 (1-677)
QY 346 AGCTTAATCTGTTTCACCTGCTTGAGGTGGTTCGAGTTCCTGCTTCCTTCCT 405
DB 114 SerPheAsnSerValHisLeuLeuAlaGlyGlyPheAlaGluPheSerArgCysPhePro 133
QY 406 GGCTCTGTAAGAAATCACTAGTCCCTACCTGATTTTCAGCTGCTGCTTACCT 465
DB 134 GlyLeuCysGlnGlyLysSerThrLeuValProThrCysIleSerGlnProCysLeuPro 153
QY 466 GTTGCACATTTGGGCAACCGCAATCTTCCCAATCTTATCTTGCTGCGAGCAGAT 525
DB 154 ValAlaAsnIleGlyProThrArgIleuProAsnLeuTyrLeuGlyCysGlnArgAsp 173
QY 526 GTTCCTCAACAG 537
DB 174 ValLeuAsnLys 177
RESULT 6
ID 099MG5 PRELIMINARY; PRT; 622 AA.
AC 099MG5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Map kinase phosphatase-M A2 isoform.

GN DUSP16 OR 3830417M17RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxId=10090;
 (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c;
 RA Matsuguchi T., Musikacharoen T., Johnson T.R., Kraft A.S.,
 RA Yoshikai Y.;
 RA "A Novel MKP is an Important Negative Regulator of LPS-mediated JNK
 RT Activation in Macrophages."
 RT Mol. Cell. Biol. 20:6999-7009(2001).
 DR EMBL: AF345952; AAK35053.1; -
 DR HSSP: Q16828; 1MKP.
 DR MGD: MGI:1917936; Dusp16.
 DR GO: GO:0005737; C:cytoplasm; IDA.
 DR GO: GO:0005634; C:nucleus; IDA.
 DR GO: GO:0005515; F:protein binding; IPT.
 DR GO: GO:000188; P:inactivation of MAPK; IDA.
 DR InterPro: IPR000340; DS phosphatase.
 DR InterPro: IPR008343; MAPK phosph.
 DR InterPro: IPR001763; Rhodanese-like.
 DR InterPro: IPR000387; TYR_phosphatase.
 DR Pfam: PF00782; DSPC; 1.
 DR Pfam: PF00581; Rhodanese; 1.
 DR PRINTS: PRO1764; MAPKPHPTASE.
 DR SMART: SM00450; RHOD; 1.
 DR SMART: SM00195; DSPC; 1.
 DR PROSITE: PS50206; RHODANSE 3; 1.
 DR PROSITE: PS00383; TYR_PHOSPHATASE 1; 1.
 DR PROSITE: PS50056; TYR_PHOSPHATASE 2; 1.
 DR PROSITE: PS50054; TYR_PHOSPHATASE_DUAL; 1.
 KW Hydrolase; kinase.
 SQ SEQUENCE 622 AA; 68672 MW; 6C0CA4EE6909B98 CRC64;

Alignment Scores:
 Pred. No.: 1.12e-40 Length: 622
 Score: 48.00 Matches: 48
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 7.20% Indels: 0
 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x Q99MG5 (1-622)

OY 586 ACCTGTCGCAAGCCTGACTTATCCCGAGTCTCATTTCTGCGTGTGCTGGAATGAC 645
 |||
 DB 139 Thrcysprolyspromaspheilleprogluserhisphleuargvalprovalasnap 158

OY 646 AGCTTTTGTGAGAAATTTTGGCGTGTGACAAATCAGTATTTTCATTGAGAAAGCA 705
 |||
 DB 159 Serphecysgilylileleuprottypleuaspyssevalaapheillegilyala 178

OY 706 AAAGCCTCAATGATGATGTTCTA 729
 |||
 DB 179 Lysalaseramglycyvalleu 186

RESULT 7
 086S88 PRELIMINARY; PRT; 625 AA.
 AC 086S88;
 DT 01-UN-2003 (TREMBLrel. 24, Created)
 DT 01-UN-2003 (TREMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Similar to dual specificity phosphatase 8.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NCBI_TaxId=9606;
 OK NCB1_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain, and Astrocytoma;

RA Strauberg R.;
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC045110; AAK45110.1; -
 DR GO: GO:0017017; F:MAP kinase phosphatase activity; IEA.
 DR GO: GO:0006470; F:protein amino acid dephosphorylation; IEA.
 DR InterPro: IPR000340; DS phosphatase.
 DR InterPro: IPR008343; MAPK phosph.
 DR InterPro: IPR002965; P-rich exten.
 DR InterPro: IPR001763; Rhodanese-like.
 DR InterPro: IPR00387; TYR_phosphatase.
 DR Pfam: PF00782; DSPC; 1.
 DR PRINTS: PRO1764; MAPKPHPTASE.
 DR PRINTS: PRO1217; PRICHEXTENSN.
 DR SMART: SM00195; DSPC; 1.
 DR SMART: SM00450; RHOD; 1.
 DR PROSITE: PS50206; RHODANSE 3; 1.
 DR PROSITE: PS00383; TYR_PHOSPHATASE 1; 1.
 DR PROSITE: PS50056; TYR_PHOSPHATASE 2; 1.
 DR PROSITE: PS50054; TYR_PHOSPHATASE_DUAL; 1.
 SQ SEQUENCE 625 AA; 65826 MW; C7C808407B724FPC CRC64;

Alignment Scores:
 Pred. No.: 3.24e-11 Length: 625
 Score: 20.00 Matches: 20
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.00% Indels: 0
 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x Q86S88 (1-625)

OY 730 GTGACGTTTGTGAGTGTGATCTCCGCTCCGACCAATGCTATGCTTACATCATGAG 789
 |||
 DB 244 ValhscysleuAlaGlyIleSerArgserAlaThrIleAlaIleAlaIleAlaIleAlaIle 263

RESULT 8
 07TS29 PRELIMINARY; PRT; 665 AA.
 ID 07TS29;
 AC 07TS29;
 DT 01-OCT-2003 (TREMBLrel. 25, Created)
 DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Dusp8 protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxId=10090;
 OK NCB1_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Brain;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hong L.,
 RA Diatchenko L., Martins K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavante T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carrincci P., Prange C.,
 RA Raba S.S., Lonnellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahney J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Buterfield Y.S.,
 RA Krzywinski M.I., Skalska U., Smalls D.E., Scherch A., Schein J.E.,
 RA Jones S.J., Watra M.A.,
 RA "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [2]
 RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6; TISSUE=Brain;
 RA Strausberg R.;
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC052705; AAF52705.1; -
 SQ SEQUENCE 665 AA; 69021 MW; 9166E36A8835249F CRC64;

Alignment Scores:
 Pred. No.: 3, 21e-11 Length: 665
 Score: 20.00 Matches: 20
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.00% Indels: 0
 DB: 11 Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x Q7NSZ9 (1-665)

QY 730 GTCACCTGTTTACCTGGATCTCCCGCTCCGCCACCATGCTATCGCTACATGAG 789
 |||||
 DB 244 ValHicylebunhagilylieserArgeralathrialeaileatylilemetlys 263

RESULT 9
 QYHV8 PRELIMINARY; PRT; 476 AA.

AC Q9YHV8;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE PUC protein (SD08157P).
 GN PUC OR CG7850.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NX NCBI_TaxID=7227;
 RX SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Manamides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champé M., Pfeiffer B.D.,
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
 RA Jaitani M., Kalish F., Karpen G.H., Ke Z., Kemison J.A., Ketchum K.A.,
 RA Juelich B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lascko P., Lai Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mikhlin N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheefer F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Slater E., Spadling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasearan D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhan S., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nuno J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.,
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AEO03677; AAF54191.1; -
 DR EMBL: AY061616; AAL29164.1; -
 DR HSSP: Q16828; 1MKP.
 DR FlyBase: Fggn0004210; puc.
 DR GO: GO:0008779; F:JUN kinase phosphatase activity; IMP.
 DR GO: GO:0017017; F:MAP kinase phosphatase activity; NAS.
 DR GO: GO:0030036; P:actin cytoskeleton organization and biogenesis; IMP.
 DR GO: GO:0046843; P:dorsal appendage formation; IMP.
 DR GO: GO:0008544; P:epidermal differentiation; IMP.
 DR GO: GO:007254; P:JNK cascade; IMP.
 DR GO: GO:0046844; P:microtubule formation; IMP.
 DR GO: GO:0046329; P:negative regulation of JNK cascade; NAS.
 DR GO: GO:0016318; P:maternal rotation; NAS.
 DR GO: GO:0006470; P:protein amino acid dephosphorylation; IDA.
 DR GO: GO:0007396; P:closure of dorsal opening; NAS.
 DR GO: GO:0042060; P:wound healing; IMP.
 DR InterPro: IPR000340; DS_phosphatase.
 DR InterPro: IPR000387; TYR_phosphatase.
 DR Pfam: PF00782; DSPC; 1.
 DR SMART: SM00195; DSPC; 1.
 DR PROSITE: PS00183; TYR_PHOSPHATASE_1; 1.
 DR PROSITE: PS00056; TYR_PHOSPHATASE_2; 1.
 DR PROSITE: PS00054; TYR_PHOSPHATASE_DUAL; 1.
 KW Hydrolyase.
 SQ SEQUENCE 476 AA; 51270 MW; 8E5DC281349F0F4F CRC64;

Alignment Scores:
 Pred. No.: 0, 0.00778 Length: 476
 Score: 13.00 Matches: 13
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.95% Indels: 0
 DB: 5 Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x Q9YHV8 (1-476)

QY 742 GCTGGATCTCCCGCTCCGCCACCATGCTATCGCTAC 780
 |||||
 DB 218 AlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyr 230

RESULT 10
 ID 046122
 AC 046122; PRELIMINARY; PRT; 476 AA.
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE PUCKERED protein.
 GN PUC OR PUCKERED OR CG7850.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NX NCBI_TaxID=7227;
 RX SEQUENCE FROM N.A.
 RA Martin-Biano E., Gampel A., Ring J., Vitdee K., Kirov N.,
 RA Tolkovsky A.M., Martinez-Arias A.,
 RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: A0223360; CAA11282.1; -
 DR HSSP: Q16828; 1MKP.
 DR FlyBase: Fggn0004210; puc.

DR GO: GO:0008579; F:JUN kinase phosphatase activity; IMP.
 DR GO: GO:0017017; F:MAP kinase phosphatase activity; NAS.
 DR GO: GO:0010036; P:actin cytoskeleton organization and biogenesis; IMP.
 DR GO: GO:0046843; P:dorsal appendage formation; IMP.
 DR GO: GO:0008544; P:epidermal differentiation; IMP.
 DR GO: GO:0007254; P:JNK cascade; IMP.
 DR GO: GO:0046844; P:micropyle formation; IMP.
 DR GO: GO:0046329; P:negative regulation of JNK cascade; NAS.
 DR GO: GO:0016318; P:ommatidial rotation; NAS.
 DR GO: GO:0006470; P:protein amino acid dephosphorylation; IDA.
 DR GO: GO:0007396; P:suture of dorsal opening; NAS.
 DR GO: GO:0042060; P:wound healing; IEA.
 DR InterPro: IPR000340; DS phosphatase.
 DR InterPro: IPR000387; TYR_phosphatase.
 DR Pfam: PF00782; DSpC; 1.
 DR SMART: SM00195; DSpC; 1.
 DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE: PS50056; TYR_PHOSPHATASE_2; 1.
 DR PROSITE: PS50054; TYR_PHOSPHATASE_DUAL; 1.
 SQ SEQUENCE 476 AA; 51269 MW; D658A08451562EC2 CRC64;

Alignment Scores:

Pred. No.:	0.000778	Length:	476
Score:	13.00	Matches:	13
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.95%	Indels:	0
DB:	5	Gaps:	0

US-10-029-345A-108_COPY_532_2532 (1-2001) x O46122 (1-476)

QY 742 GCTGGATCTCCGCTCCGCCACCATCGCTATCGCTTAC 780
 |||||
 Db 218 ALaGlyIleSerArgSerAlaThrIleAlaIleAlaIleTyr 230

RESULT 11

Q82ZQ0 PRELIMINARY; PRT; 498 AA.
 AC Q82ZQ0;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DE Cytosine/purines, uracil, thiamine, allantoin permease family
 DE protein.
 GN EF3000.
 OS Enterococcus faecalis (Streptococcus faecalis).
 OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
 OX NCBI_TaxID=1351;
 RN 11)
 RP SEQUENCE FROM N.A.
 RC STRAIN=V583 / ATCC 700802;
 RX MEDLINE=22550857; PubMed=12663927;
 RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seshadri R.,
 Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,
 Tettelin H., Dodson R.J., Umayam L., Brinkay L., Beanan M.,
 Daugherty S., DeBoy R.T., Durkin S., Kolonay J., Maduprat R., Nelson W.,
 Vamathevan J., Tian B., Upton J., Hansen T., Shetty J., Khouli H.,
 Ullrich T., Raddue D., Ketchum K.A., Dougherty B.A., Fraser C.M.;
 RT "Role of mobile DNA in the evolution of vancomycin-resistant
 Enterococcus faecalis.";
 RL Science 299:2071-2074 (2003).
 DR EMBL: AE016956; AAO82685.1; -.
 DR TIGR: EF3000; -.
 DR GO: GO:0016020; C:membrane; IEA.
 DR GO: GO:0015205; F:nucleobase transporter activity; IEA.
 DR GO: GO:0015931; F:nucleobase, nucleoside, nucleotide and nucl. .; IEA.
 DR InterPro: IPR001248; Cyt_pur_permease.
 DR Pfam: PF02133; Transp_cyt_pur; 1.
 SQ SEQUENCE 498 AA; 54165 MW; FF318187B7B3C90 CRC64;

Alignment Scores: 1.11 Length: 498
 Pred. No.:

Score:	10.00	Matches:	10
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.55%	Indels:	0
DB:	16	Gaps:	0

US-10-029-345A-108_COPY_532_2532 (1-2001) x Q82ZQ0 (1-498)

QY 103 GCCGCTATCATTTAGCAGCACTTTTCG 74
 |||||
 Db 342 ALaGlyTyrGlnLeuAlaAlaLeuPhePro 351

RESULT 12

Q89P10 PRELIMINARY; PRT; 550 AA.
 AC Q89P10;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE BLR3500 protein.
 GN BLR3500.
 OS Bradyrhizobium japonicum.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Bradyrhizobiaceae; Bradyrhizobium.
 OX NCBI_TaxID=375;
 RN 11)
 RP SEQUENCE FROM N.A.
 RC STRAIN=USDA 110;
 RX MEDLINE=2248498; PubMed=12597275;
 RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
 Sasamoto S., Matanabe A., Ideawa K., Iriyuchi M., Kawashima K.,
 Kohara M., Matsunoto M., Shimpo S., Tsunoka H., Wada T., Yamada M.,
 Tabeta S.;
 RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
 Bradyrhizobium japonicum USDA110.";
 RL DNA Res. 9:189-197 (2002).
 DR EMBL: AP005947; BAC48765.1; -.
 DR GO: GO:0016020; C:membrane; IEA.
 DR GO: GO:0009306; P:protein secretion; IEA.
 DR InterPro: IPR001775; Bac_GSPD.
 DR InterPro: IPR004846; GSP1/IIIProtein.
 DR Pfam: PF00263; Secretin; 1.
 DR PRINTS: PR00811; BCTERIALGSPD.
 KW Complete proteome.
 SQ SEQUENCE 550 AA; 56757 MW; 185922C640A1782A CRC64;

Alignment Scores:

Pred. No.:	12.3	Length:	550
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.40%	Indels:	0
DB:	16	Gaps:	0

US-10-029-345A-108_COPY_532_2532 (1-2001) x Q89P10 (1-550)

QY 1100 CTGGCAGCTGCGGACAGCTGGCGGA 1074
 |||||
 Db 269 LeuGlyThrLeuGlyThrLeuAlaGly 277

RESULT 13

Q88T18 PRELIMINARY; PRT; 606 AA.
 AC Q88T18;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Hypothetical protein F08B1.1b.
 GN F08B1.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Chordata; Rhabdilitida; Rhabdilitidea;
 OC Rhabdilitidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Briscot N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
  investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Briscot N2;
RA Chisee S.;
RT "The sequence of C. elegans coamid F08B1.";
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Briscot N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U23178; AAK68300.1; -.
DR HSSP; Q16828; IMKP.
DR WormBep; F08B1.1b; CE27919.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0008138; F:protein tyrosine/serine/threonine phosphata. . .; IEA.
DR GO; GO:0006470; P:protein amino acid dephosphorylation; IEA.
DR InterPro; IPR000340; DS_phosphatase.
DR InterPro; IPR000387; TYR_phosphatase.
DR Pfam; PF00782; DSPC; 1.
DR SMART; SM00195; DSPC; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS50054; TYR_PHOSPHATASE_DUAL; 1.
DR Hypothetical protein; Hydrolase.
SQ SEQUENCE 606 AA; 64962 MW; 5B71795C6ED58058 CRC64;

Alignment Scores:
Pred. No.: 12 2 Length: 606
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.35% Indels: 0
DB: 5 Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x Q8ST18 (1-606)

QY 733 CACTGTTAGCTGGATCTCCGCTCC 759
Db ||||| 210 HisCysLeuAlaGlyIleSerArgSer 218

RESULT 14
Q8ST19 PRELIMINARY; PRT; 657 AA.
AC Q8ST19;
DT 01-JUN-2002 (T-EMBLrel. 21, Created)
DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Hypothetical protein F08B1.1a.
DE F08B1.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditicoidea;
OC Rhabditidae; Pelodolerinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Briscot N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
  investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
```

```

RC STRAIN=Briscot N2;
RA Chisee S.;
RT "The sequence of C. elegans coamid F08B1.";
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Briscot N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U23178; AAK68299.1; -.
DR HSSP; Q16828; IMKP.
DR WormBep; F08B1.1a; CE27918.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0017017; F:MAP kinase phosphatase activity; IEA.
DR GO; GO:0006470; P:protein amino acid dephosphorylation; IEA.
DR InterPro; IPR000340; DS_phosphatase.
DR InterPro; IPR008343; MAPK_phosph.
DR InterPro; IPR001763; Rhodanese-like.
DR InterPro; IPR000387; TYR_phosphatase.
DR Pfam; PF00782; DSPC; 1.
DR PRINTS; PR01764; MAPKPHPTASE.
DR SMART; SM00195; DSPC; 1.
DR SMART; SM00450; RHODANSE; 1.
DR PROSITE; PS50206; RHODANSE_3; 1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS50054; TYR_PHOSPHATASE_DUAL; 1.
DR PROSITE; PS50054; TYR_PHOSPHATASE_DUAL; 1.
DR Hypothetical protein; Hydrolase.
SQ SEQUENCE 657 AA; 71002 MW; A0D9153DE6326B43 CRC64;

Alignment Scores:
Pred. No.: 12 Length: 657
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.35% Indels: 0
DB: 5 Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x Q8ST19 (1-657)

QY 733 CACTGTTAGCTGGATCTCCGCTCC 759
Db ||||| 261 HisCysLeuAlaGlyIleSerArgSer 269

RESULT 15
Q8PEN9 PRELIMINARY; PRT; 486 AA.
AC Q8PEN9;
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Outer membrane efflux protein.
DE NODT OR XAC4306.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
Camarote G., Cannavan F., Cardoso J., Chamberg F., Ciapina L.P.,
Ciccarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
Lacali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rosel N.M.,
Martins E.C., Meldans J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
```

RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Teiza R.I.D.,
RA Spradode dos Santos M., Triffi D., Tsai S.M., White F.F.,
RA Serubal J.C., Kitajima J.P.,
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities";
RL Nature 417:459-463 (2002).
DR EMBL; AB012084; AAM39136.1; -
DR GO; GO:0005215; P:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003423; OEP; Prok_lipoprot_S.
DR InterPro; IPR004377; Prok_lipoprot_S.
DR Pfam; PF02321; OEP; 2.
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Complete proteome.
SQ SEQUENCE 486 AA; 51711 MW; 7E0D7E71A18F02D CRC64;

Alignment Scores:

Pred. No.: 141 Length: 486
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0
DB: 16 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x Q8PEN9 (1-486)

QY 103 GCCGGTATCATTTAGCAGCACTT 80

DB 253 AAGGlyYrGlnIneuAlaAlaLeu 260

RESULT 16

ID Q9H6T2 PRELIMINARY; PRT; 495 AA.

AC Q9H6T2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ21915.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obyashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEO human cDNA sequencing project";
RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK025568; BAB15171.1; -
DR GO; GO:0003899; P:DNA-directed RNA polymerase activity; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR InterPro; IPR007081; RNA_pol_Rpb1_5.
DR Pfam; PF04998; RNA_pol_Rpb1_5; 1.
KW Hypothetical protein.
SQ SEQUENCE 495 AA; 56749 MW; F30E6916AD9BED3 CRC64;

Alignment Scores:

Pred. No.: 141 Length: 495
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 4 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x Q9H6T2 (1-495)

QY 386 AGTCTCTGTTGTTTCCCTGACC 409

DB 316 SerSerLeuValValSerLeuAla 323

RESULT 17

Q96AG9
ID Q96AG9 PRELIMINARY; PRT; 495 AA.
AC Q96AG9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Tissue-Skin;
RA Strausberg R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC017115; AAH17115.1; -
DR GO; GO:0003899; P:DNA-directed RNA polymerase activity; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR InterPro; IPR007081; RNA_pol_Rpb1_5.
DR Pfam; PF04998; RNA_pol_Rpb1_5; 1.
KW Hypothetical protein.
SQ SEQUENCE 495 AA; 56750 MW; 2D7AE1D91E5D7871 CRC64;

Alignment Scores:

Pred. No.: 141 Length: 495
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 4 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x Q96AG9 (1-495)

QY 386 AGTCTCTGTTGTTTCCCTGACC 409

DB 316 SerSerLeuValValSerLeuAla 323

RESULT 18

ID Q88T27 PRELIMINARY; PRT; 497 AA.

AC Q88T27;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Threonine synthase (EC 4.2.99.2).
GN THRC OR LP 2758.
OS Lactobacillus plantarum.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OC NCBI_TaxID=1590;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=NCIMB 8826 / WCPS1;
RX MEDLINE=22480296; PubMed=12566566;
RA Kleerebezem M., Boekhorst J., van Kranenburg R., Molenaar D.,
RA Kuipers O.P., Leer R., Turchini R., Peters S.A., Sandbrink H.M.,
RA Fiers M.W.E.J., Stiekema W., Klein Lankhorst R.M., Bron P.A.,
RA Hofler S.M., Nierop Groot W.N., Kerkhoven R., De Vries M., Ureing B.,
RA De Vos W.M., Siezen R.J.;
RT "Complete genome sequence of Lactobacillus plantarum WCPS1";
RL Proc. Natl. Acad. Sci. U.S.A. 100:1990-1995 (2003).
DR EMBL; AL935260; CAD64989.1; -
DR GO; GO:0016829; P:lyase activity; IEA.
DR GO; GO:0006520; P:amino acid metabolism; IEA.
DR InterPro; IPR001926; B6_enzyme_beta.
DR InterPro; IPR000634; S/T_dehydrtse_BS.
DR Pfam; PF00291; PALP; 1.
DR PROSITE; PS00165; DEHYDRATASE_SER_THR; 1.
KW Lyase; Complete proteome.
SQ SEQUENCE 497 AA; 54569 MW; DA4344785D0275CC CRC64;

Alignment Scores:

Pred. No.: 141
 Score: 8.00
 Percent Similarity: 100.00%
 Best Local Similarity: 100.00%
 Query Match: 1.20%
 DB: 16
 Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x Q8658 (1-497)

QY 1733 GTGCCACTTACTCTGCTACAGCT 1756

DB 34 ValProValThrLeuProThrAla 41

RESULT 19

Q8658 PRELIMINARY; PRT; 513 AA.

AC Q8658; 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DE Putative transposable element.
 GN OSJNB0004A10.10.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Nipponbare;
 RA McCombie W.R., de la Bastide M., Spiegel L., Nascimento L., Ballja V.,
 RA Zuberavon T., Bell M., Preston R., Kirchoff K., Kuit K., Baker J.,
 RA Santos L., Miller B., Cumins D.M., Katzenberger F., Muller S.,
 RA Shah R., King L., Yang C., Dike S., O'Shaughnessy A., Palmer L.,
 RA Dedha N.;
 RT "Genomic sequence for Oryza sativa, Nipponbare strain, clone
 RT OSJNB0004A10, from chromosome 10, complete sequence."
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC098682; AA01106.1; -.
 DR Gramene; Q8658; -.
 DR InterPro; IPR008906; HATC.
 DR Pfam; PF05699; hATC; 1.
 SQ SEQUENCE 513 AA; 58087 MW; 978805247396E6AA CRC64;

Alignment Scores:

Pred. No.: 140 Length: 513
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.24% Indels: 0
 DB: 10 Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x Q8658 (1-513)

QY 637 CAGCACACGACGAATGAGACT 614

DB 490 GlnAlaHisAlaGlyAsnGlnThr 497

RESULT 20

Q7XEV9 PRELIMINARY; PRT; 513 AA.

AC Q7XEV9; 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DE Putative transposable element.
 GN OSJNB0004A10.10.
 OS Oryza sativa (Japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=39947;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA The Rice Chromosome 10 Sequencing Consortium;
 RT "In-depth view of structure, activity, and evolution of rice
 RT chromosome 10."
 RL Science 300:1566-1569(2003).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Buell C.R., Wang R.A., McCombie W.R., Messing J., Yuan Q.;
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB017075; AAP52942.1; -.
 SQ SEQUENCE 513 AA; 58087 MW; 978805247396E6AA CRC64;

Alignment Scores:

Pred. No.: 140 Length: 513
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.24% Indels: 0
 DB: 10 Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x Q7XEV9 (1-513)

QY 637 CAGCACACGACGAATGAGACT 614

DB 490 GlnAlaHisAlaGlyAsnGlnThr 497

RESULT 21

Q92M58 PRELIMINARY; PRT; 513 AA.

AC Q92M58; 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DE Putative sigma-54-dependent transcription regulator protein.
 GN TACA OR R02795 OR SMC04011.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RA MEDLINE=21386507; PubMed=11481430;
 RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
 RA Bisterard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
 RA Godrie T., Goffeau A., Kahn D., Kles A., Lelaure V., Masny D.,
 RA Pohl T., Portet-Luc A., Puhler A., Purnelle B., Ramsparger U.,
 RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;
 RT "Analysis of the chromosome sequence of the legume symbiont
 RT Sinorhizobium meliloti strain 1021."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
 DR EMBL; AL591791; CAC47374.1; -.
 DR HSSP; P06143; 1D4Z.
 DR GO; GO:0000166; P: nucleotide binding; IEA.
 DR GO; GO:0003700; P: transcription factor activity; IEA.
 DR GO; GO:0000156; P: two-component response regulator activity; IEA.
 DR GO; GO:0000160; P: two-component signal transduction system (p. . .); IEA.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR002197; HTH_Fig.
 DR InterPro; IPR001789; Response_reg.
 DR InterPro; IPR008298; Reg_reg_NtcC.
 DR InterPro; IPR002078; Sig54_interac.
 DR Pfam; PF02954; HTH_8; 1.
 DR Pfam; PF00072; response_reg; 1.
 DR Pfam; PF00158; Sigma54_activat; 1.
 DR PRINTS; PRO1590; HTHRS.
 DR ProDom; PD000039; Response_reg; 1.
 DR SMART; SM00382; AAA; 1.
 DR SMART; SM00448; REC; 1.
 DR TIGRFAMs; TIGR01199; HTH_fib; 1.
 DR PROSITE; PSS0110; RESPONSE_REGULATORY; 1.

DR PROSITE; PS00675; SIGMA54 INTERACT_1; 1.
 DR PROSITE; PS00676; SIGMA54 INTERACT_2; 1.
 DR PROSITE; PS00688; SIGMA54 INTERACT_3; 1.
 DR PROSITE; PS50045; SIGMA54 INTERACT_4; 1.
 DR PIRSF; PIRSF003187; Res_Reg_Ntrc; 1.
 KW Complete proteome.
 SQ SEQUENCE 513 AA; 55902 MW; 5E1964533DE8F65 CRC64;

Alignment Scores:

Pred. No.:	140	Length:	513
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.24%	Indels:	0
DB:	16	Gaps:	0

US-10-029-345A-108_COPY_532_2532 (1-2001) x Q92M58 (1-513)

QY 1776 GCAAGTGGGACGCTGGCTGCACT 1753

Db 417 AlaSerGlyGlnLeuAlaAla 424

RESULT 22

Q9FX19 PRELIMINARY; PRT; 515 AA.

AC Q9FX19; 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE F6F9.1 Protein (Altig19940/F6F9_1).
 GN F6F9.1.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 NCBI_TaxID=3702;
 [1]
 RP SEQUENCE FROM N.A.

RA Alferf N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
 RA Alefati H., Nguyen M., Lam B., Southwick A., Miranda M., Brooks S.,
 RA Buchler E., Chao Q., Chin C., Chou J., Choi E., Gonzalez A.,
 RA Hwang B., Johnson-Hopson C., Khan S., Kim C., Koo T., Lee J.M.,
 RA Lenz C., Liu A., Liu S., Mukharly N., Pham P., Sakano H., Shinn P.,
 RA Toriumi M., Vayberg M., Yu G., Ecker J., Theologis A., Davis R.W.,
 RL Submitted (SFP-2000) to the EMBL/GenBank/DBJ databases.
 [2]
 RP SEQUENCE FROM N.A.

RA Shinn P., Chen H., Cheuk R., Kim C.J., Koesema B., Meyers M.C.,
 RA Banh J., Bowser L., Carninci P., Dale J.M., Gibson H.A.,
 RA Goldsmith A.D., Hayashizaki Y., Ishida J., Jiang P.X., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J.,
 RA Liu S.X., Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,
 RA Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
 RA Tang C.C., Toriumi M., Yamada K., Yu G., Yu S., Shinozaki K.,
 RA Davis R.W., Theologis A., Ecker J.R.,
 RT "Arabidopsis cDNA clones."
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 [3]
 RP SEQUENCE FROM N.A.

RA Kim C.J., Chen H., Cheuk R., Shinn P., Banh J., Bowser L.,
 RA Carninci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
 RA Ishida J., Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Lam B.,
 RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,
 RA Palm C.J., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
 RA Tang C.C., Toriumi M., Wu H.C., Yamada K., Yamamura Y., Yu G., Yu S.,
 RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.,
 RT "Arabidopsis ORF clones."
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC007797; AAG12562.1; -
 DR EMBL; AY048245; AKK82507.1; -
 DR EMBL; AY13063; AAM47371.1; -
 DR PIR; G86332; G86332.
 DR HSP; P26221; ITF4.

DR GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . .; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR InterPro; IPR001701; Glyco_hydro_9.
 DR InterPro; IPR008928; Glyco_trans_6np.
 DR Pfam; PF00759; Glyco_hydro_9; 1.
 DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
 SQ SEQUENCE 515 AA; 56707 MW; F21BEC7479B5E209 CRC64;

Alignment Scores:

Pred. No.:	140	Length:	515
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.20%	Indels:	0
DB:	10	Gaps:	0

US-10-029-345A-108_COPY_532_2532 (1-2001) x Q9FX19 (1-515)

QY 374 GTGGGTTTCTGAGTTCTTCGTT 397

Db 499 ValGlyLeuLeuSerSerLeuVal 506

RESULT 23

Q8LCP6 PRELIMINARY; PRT; 525 AA.

AC Q8LCP6; 01-OCT-2002 (TREMBLrel. 22, Created)
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
 DE Endo-beta-1,4-glucanase, putative.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 NCBI_TaxID=3702;
 [1]
 RP SEQUENCE FROM N.A.

RA Haas B.J., Volkovsky N., Town C.D., Troughan M., Alexandrov N.,
 RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.,
 RT "Full-length messenger RNA sequences greatly improve genome
 annotation.";
 RL Genome Biol. 0:0-0(2002).
 [2]
 RP SEQUENCE FROM N.A.

RA Brover V., Troughan M., Alexandrov N., Lu Y.-P., Flavell R.,
 RT "Full-length cDNA from Arabidopsis thaliana."
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY086475; AAM63477.1; -
 DR GO; GO:0004553; F:hydrolyase activity, hydrolyzing O-glycosyl . . .; IEA.
 DR GO; GO:0005975; P:carbohydrate metabolism; IEA.
 DR InterPro; IPR001701; Glyco_hydro_9.
 DR InterPro; IPR008928; Glyco_trans_6np.
 DR Pfam; PF00759; Glyco_hydro_9; 1.
 DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
 SQ SEQUENCE 525 AA; 57915 MW; 26D724334F5E32A CRC64;

Alignment Scores:

Pred. No.:	140	Length:	525
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.20%	Indels:	0
DB:	10	Gaps:	0

US-10-029-345A-108_COPY_532_2532 (1-2001) x Q8LCP6 (1-525)

QY 374 GTGGGTTTCTGAGTTCTTCGTT 397

Db 508 ValGlyLeuLeuSerSerLeuVal 515

RESULT 24

Q9LR07

ID Q9LR07 PRELIMINARY; PRT; 525 AA.
AC Q9LR07;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE F10A5.13 (Hypothetical protein) (Putative
endo-beta-1,4-glucanase).
GN F10A5.13 OR ATIG75680.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_Taxid=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA Khan S., Kim C., Altafli H., Bei B., Chin C., Chlou J., Choi E.,
RA Conn L., Conway A., Gonzalez A., Hansen N., Howling B., Koo T., Lam B.,
RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukhareky N.,
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA Thaveri A., Toriumi M., Vayberg M., Yu G., Davis R., Federspiel N.,
RA Theologis A., Ecker J.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA Khan S., Kim C., Altafli H., Bei B., Chin C., Chlou J., Choi E.,
RA Conn L., Conway A., Gonzalez A., Hansen N., Howling B., Koo T., Lam B.,
RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukhareky N.,
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA Thaveri A., Toriumi M., Vayberg M., Yu G., Davis R., Federspiel N.,
RA Theologis A., Ecker J.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Dale J.M., Gibson H.A., Goldemich A.D., Jiang P.X., Lee J.M.,
RA Quach H.L., Tang C., Toriumi M., Yu G., Bowser L., Carninci P.,
RA Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T., Kaniya A.,
RA Karlin-Newman G., Kawai J., Kim C., Koesema E., Lam B., Lin J.,
RA Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E.,
RA Shinzaki K., Davis R.W., Ecker J.R., Theologis A.;
RL "Full length cDNA of gene F10A5.13 (GI9359363).";
RT Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Hansen V.W., Lee J.M.,
RA Quach H.L., Tang C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,
RA Yu G., Yuan S., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,
RA Ishida J., Jones T., Kaniya A., Kawai J., Kim C.J., Narusaka M.,
RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,
RA Southwick A., Tripp M.G., Wu T., Shinzaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RT "Arabidopsis Open Reading Frame (ORF) Clones.";
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN EMBL; AC006434; AA087112.1; -
DR EMBL; AY039938; AA064042.1; -
DR EMBL; AY150451; AA012892.1; -
DR PIR; E96786; E96786.
DR HSSP; P26221; 1TF4.
DR GO; GO:0004553; F:hydrolase activity, hydrolyzing O-glycosyl . . .; IEA.
DR GO; GO:0005755; P:carbohydrate metabolism; IEA.
DR InterPro; IPR001701; Glyco_hydro_9.
DR InterPro; IPR008928; Glyco_trans_6p.
DR Pfam; PF00759; Glyco_hydro_9; 1.
DR PROSITE; PS00592; GLYCOSYL_HYDROL_F9_1; 1.
KW Hypothetical protein.
SQ SEQUENCE 525 AA; 57868 MW; FBC3704F4BB8B71 CRC64;

Alignment Scores: 140 Length: 525
Pred. No.: 8.00 Matches: 0
Score: 8.00

Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 10 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x Q9LR07 (1-525)

QY 374 GTGGCTTGCTGAGTCTCTCGTT 397
Db 508 ValGlyLeuSerSerLeuVal 515

RESULT 25

ID Q65756 PRELIMINARY; PRT; 526 AA.

AC Q65756;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
OS VP5.
OC Blueongue virus.
OC Viruses; dsRNA viruses; Reoviridae; Orbivirinae.
OX NCBI_Taxid=40051;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=serotype 11 strain UC-8;
RT "Association of Blueongue Virus Gene Segment 5 with Neurovirulence.";
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
RN EMBL; U03285; AAA03474.1; -
DR GO; GO:0019028; C:Viral capsid; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000145; Orbi_VP5.
DR Pfam; PF0901; Orbi_VP5; 1.
SQ SEQUENCE 526 AA; 59273 MW; D46136DDA1174586 CRC64;

Alignment Scores:

Pred. No.: 140 Length: 526
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0
DB: 12 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x Q65756 (1-526)

QY 98 CTATCAATTAGACGACTTTTCC 75
Db 435 LeuSerLeuSerSerThrPheSer 442

RESULT 26

ID Q9Y0X9 PRELIMINARY; PRT; 534 AA.

AC Q9Y0X9;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DE 01-MAY-2003 (TrEMBLrel. 23, Last annotation update)
OS VP5.
GN COS54_23.
GN Rand herpesvirus 1 (lucke tumor herpesvirus).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae.
OX NCBI_Taxid=85655;
RN [1]
RP SEQUENCE FROM N.A.
RA Davison A., Sauerbier W., Dolan A., Addison C., McKinnell R.G.,
RT "Genomic studies of the Lucke tumor herpesvirus (RaHV-1).";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN EMBL; A011004; A0112286.1; -
DR InterPro; IPR003006; IG_MHC.
DR PROSITE; PS00290; IG_MHC; 1.
SQ SEQUENCE 534 AA; 59141 MW; 0873A74AC0ACCECF9 CRC64;

Alignment Scores: 140 Length: 534
Pred. No.: 8.00 Matches: 0
Score: 8.00

Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.20% Indels: 0
 DB: 12 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x Q9YQX9 (1-534)

QY 1092 CGTGCCGAGCGTGGAGCCGCGCT 1115
 DB 367 ArgAlaGlnArgAlaAlaValAla 374

RESULT 27

ID Q82GP4 PRELIMINARY; PRT; 544 AA.

AC Q82GP4; 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DE Putative integral membrane transport protein.
 GN SAV3853.
 OS Streptomyces avermitilis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 CC Streptomycinae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=33903;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=21477403; PubMed=11572948;
 RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
 RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
 RA Kikuchi H., Shiba T., Sakaki Y., Hatori M.;
 RT "Genome sequence of an industrial microorganism Streptomyces
 RT avermitilis: deducing the ability of producing secondary
 RT metabolites.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=22608306; PubMed=12692562;
 RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
 RA Sakaki Y., Hatori M., Omura S.;
 RT "Complete genome sequence and comparative analysis of the industrial
 RT microorganism Streptomyces avermitilis.";
 RL Nat. Biotechnol. 21:526-531(2003).
 DR EMBL; AP005036; BAC1555.1; -.
 KW Complete proteome.
 SQ SEQUENCE 544 AA; 56291 MW; FC13A5045264A9DA CRC64;

Alignment Scores:
 Pred. No.: 139 Length: 544
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.20% Indels: 0
 DB: 16 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x Q82GP4 (1-544)

QY 1648 ACCTTACCCCTTCCTGACGAC 1671

DB 13 ThnSerThProSerLeuThrSer 20

RESULT 28

ID Q92ZH5 PRELIMINARY; PRT; 545 AA.

AC Q92ZH5; 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DE Putative arylsulfatase.
 GN RA0510 OR SMA0943.
 OS Rhizobium meliloti (Sinorhizobium meliloti).

OG Plasmid pSymA (megaplasmid 1).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 CC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=21396509; PubMed=11481432;
 RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
 RA Barlow-Hubler F., Bowser L., Capela D., Gilbert F., Gouy J.,
 RA Gurjal M., Hong A., Hutzler L., Hyman R.W., Kahn D., Kahn M.L.,
 RA Kaiman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
 RA Yen K.-C., Davis R.W., Federspiel N.A., Long S.R.;
 RT "Nucleotide sequence and predicted functions of the entire
 RT Sinorhizobium meliloti pSymA megaplasmid.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
 DR EMBL; AR007242; AKK65168.1; -.
 DR PIR; P95325; P95325.
 DR GO; GO:0046821; C:cyttrachromosomal DNA; IEA.
 DR GO; GO:0008484; F:sulfuric ester hydrolase activity; IEA.
 DR GO; GO:0008152; P:metabolism; IEA.
 DR InterPro; IPR000917; Sulfatase.
 DR Pfam; PF00884; Sulfatase.
 KW plasmid; Hypothetical protein; Complete proteome.
 SQ SEQUENCE 545 AA; 60458 MW; 629CC3D9A07A3C13 CRC64;

Alignment Scores:
 Pred. No.: 139 Length: 545
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.24% Indels: 0
 DB: 16 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x Q92ZH5 (1-545)

QY 1593 CGTGAGTGTGCTGCTGCTGCGA 1570

DB 15 ArgGlnValIleuValAlaGlyGly 22

RESULT 29

ID O55887 PRELIMINARY; PRT; 555 AA.

AC O55887; 01-JUN-1998 (TrEMBLrel. 06, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Fusion glycoprotein (F glycoprotein).
 GN F.

OS Human parainfluenza virus 1.
 OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
 CC Paramyxoviridae; Paramyxovirinae; Respirovirus.
 OX NCBI_TaxID=12730;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=PIV1/Washington/20993/1964;
 RX MEDLINE=75188544; PubMed=166929;
 RA Murphy B.R., Richman D.D., Chaihub E.G., Unlendorf C.P., Baron S.,
 RA Chanock R.M.;
 RT "Failure of attenuated temperature-sensitive influenza A (H3N2) virus
 RT to induce heterologous interference in humans to parainfluenza type 1
 RT virus.";
 RL Infect. Immun. 12:62-68(1975).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PIV1/Washington/20993/1964;
 RA Tao T., Durbin A.P., Whitehead S.S., Davoodi F., Collins P.L.,
 RA Murphy B.R.;
 RT "Recovery of a chimeric human parainfluenza virus from PIV3 cDNA in
 RT which the hemagglutinin and fusion glycoproteins of PIV3 were
 RT substituted for those of PIV1.";
 RN Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]

RP SEQUENCE FROM N.A.
RC STRAIN=Washington 1964;
RA Newman J.T., Surman S.R., Riggs J.M., Hansen C.T., Collins P.L.,
Murphy B.R., Skidopoulos M.H.;
RT "Sequence Analysis of the Washington/1964 Strain of Human
Parainfluenza Virus Type 1 (HPIV1) and Recovery and Characterization
of Wild Type Recombinant HPIV1 Produced by Reverse Genetics.";
RL Virus Genes 24:0-0(2002).
RN (4)
RP SEQUENCE FROM N.A.
RC STRAIN=Washington 1964;
RA Newman J.T., Surman S.R., Riggs J.M., Hansen C.T., Collins P.L.,
Murphy B.R., Skidopoulos M.H.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF016279; AAC23945.1; -
DR EMBL; AF457102; AAL89407.1; -
DR GO; GO:0019039; F:Viral-cell fusion molecule activity; IEA.
DR GO; GO:0006948; P:Viral-induced cell-cell fusion; IEA.
DR InterPro: IPR000776; Fusion.gly.
DR Pfam: PF00523; fusion.gly.1.
SQ SEQUENCE 555 AA; 60768 MW; 439F9BC2FE1D4EEF CRC64;

Alignment Scores:
Pred. No.: 139 Length: 555
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0
DB: 12 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x 055887 (1-555)
QY 1193 AGCTTATGCTGCTTCAGCCTG 1170
Db 13 SerLeuLeuSerSerSerSerLeu 20

RESULT 30
055888 PRELIMINARY; PRT; 555 AA.
ID 055888
AC 055888;
DT 01-JUN-1998 (TRENBLrel. 06, Created)
DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Fusion glycoprotein.
GN F.
OS Recombinant PIV3/PIV1 virus.
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae.
OC NCBI_TaxID=65405;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=PIV3-1;
RX MEDLINE=97428580; PubMed=9281512;
RA Durbin A.P., Hall S.L., Siew J.W., Whitehead S.S., Collins P.L.,
Murphy B.R.;
RT "Recovery of infectious human parainfluenza virus type 3 from cDNA.";
RL Virology 235:323-332(1997).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=PIV3-1;
RX MEDLINE=98184528; PubMed=9525616;
RA Tao T., Durbin A.P., Whitehead S.S., Davoodi F., Collins P.L.,
Murphy B.R.;
RT "Recovery of a fully viable chimeric human parainfluenza virus (PIV)
type 3 in which the hemagglutinin-neuraminidase and fusion
glycoproteins have been replaced by those of PIV type 1.";
RL J.Virol. 77:2955-2961(1998).
DR EMBL; AF016281; AAC23947.1; -
DR GO; GO:0019039; F:Viral-cell fusion molecule activity; IEA.
DR GO; GO:0006948; P:Viral-induced cell-cell fusion; IEA.
DR InterPro: IPR000776; Fusion.gly.
DR Pfam: PF00523; fusion.gly.1.
SQ SEQUENCE 555 AA; 60768 MW; 439834D88055CAEF CRC64;

Alignment Scores:
Pred. No.: 139 Length: 555
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0
DB: 12 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x 055888 (1-555)
QY 1193 AGCTTATGCTGCTTCAGCCTG 1170
Db 13 SerLeuLeuSerSerSerSerLeu 20

RESULT 31
08ABA4 PRELIMINARY; PRT; 559 AA.
ID 08ABA4
AC 08ABA4;
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Asparagine synthetase B.
GN BT0551.
OS Bacteroides thetaiotaomicron.
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Bacteroidaceae; Bacteroides.
OC NCBI_TaxID=818;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=VPI-5482 / ATCC 29148;
RX MEDLINE=22550858; PubMed=12663928;
RA Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,
Chiang H.C., Hooper L.V., Gordon J.I.;
RT "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis.";
RL Science 299:2074-2076(2003).
DR EMBL; AE016928; MA07658.1; -
DR GO; GO:0004066; P:asparagine synthase (glutamine-hydrolyzing) . . .; IEA.
DR GO; GO:0006529; P:asparagine biosynthesis; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro: IPR001962; Asn_synthase.
DR InterPro: IPR006426; Asn_synth_AEB.
DR InterPro: IPR000583; GATase_2.
DR Pfam: PF00733; Asn_synthase; 1.
DR Pfam: PF00310; GATase_2; 1.
DR TIGRPFam; TIGR01536; asn_synth_AEB; 1.
KM Complete proteome.
SQ SEQUENCE 559 AA; 63019 MW; 292F181736AD4B05 CRC64;

Alignment Scores:
Pred. No.: 139 Length: 559
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 16 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x 08ABA4 (1-559)
QY 1081 AGCGTCCAGCGTGCCAGCGTG 1104
Db 514 SerValProSerValProSerVal 521

RESULT 32
07MG72 PRELIMINARY; PRT; 569 AA.
ID 07MG72
AC 07MG72;
DT 01-OCT-2003 (TRENBLrel. 25, Created)
DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Putative oxidoreductase.
GN BB4047.
OS Bordetella bronchiseptica (Alcaligenes bronchiseptica).

OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
RN [1]_TaxID=518;
RP SEQUENCE FROM N.A.
RC STRAIN=RB50 / ATCC BAA-568;
RX MEDLINE=22827954; PubMed=12910271;
RA Parikhil J., Sebakhia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.M., Temple L., James K., Harris B., Cherevach I.,
RA Achman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leathers S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Umwin L., Whitehead S., Barrell B.G., Maskell D.J.,
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.",
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX64049; CAE34410.1; -.
KW Complete proteome.
SQ SEQUENCE 569 AA; 61713 MW; DAF4235C8D86A5BE CRC64;

Alignment Scores:
Pred. No.: 138 Length: 569
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0
DB: 16 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x Q7W672 (1-569)

OY 1022 GGACTGAGGGCGTCTCGCTTTTC 999
DB 241 GlyleuarglyValserLeuPhe 248

RESULT 33

Q7W4P6 PRELIMINARY; PRT; 569 AA.

AC Q7W4P6; 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DE Putative oxidoreductase.
GN BP3612
OS Bordetella parapertussis.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=519;
RN [1]_TaxID=519;
RP SEQUENCE FROM N.A.
RC STRAIN=12822 / ATCC BAA-587;
RX MEDLINE=22827954; PubMed=12910271;
RA Parikhil J., Sebakhia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdano-Tarraga A.M., Temple L., James K., Harris B., Cherevach I.,
RA Achman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leathers S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Umwin L., Whitehead S., Barrell B.G., Maskell D.J.,
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.",
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX64043; CAE38896.1; -.
KW Complete proteome.
SQ SEQUENCE 569 AA; 61714 MW; D3D9B5B5B72A0DC CRC64;

Alignment Scores:

Pred. No.: 138 Length: 569
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0
DB: 16 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x Q7W4P6 (1-569)

OY 1022 GGACTGAGGGCGTCTCGCTTTTC 999
DB 241 GlyleuarglyValserLeuPhe 248

RESULT 34

Q7U0D7 PRELIMINARY; PRT; 573 AA.

AC Q7U0D7; 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DE Periplasmic glucan biosynthesis protein Mdog.
GN MDOG OR RB6383.
OS Rhodopirellula baltica.
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC Planctomycetaceae; Pirellula.
OX NCBI_TaxID=117;
RN [1]_TaxID=117;
RP SEQUENCE FROM N.A.
RC STRAIN=1;
RX MEDLINE=22735913; PubMed=12835416;
RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Borzym K., Hettmann K., Rabus R.,
RA Schlesner H., Amann R., Reinhardt R.,
RT "Complete genome sequence of the marine planctomycete Pirellula sp.
RT strain 1.",
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303 (2003).
DR EMBL; BX294144; CAD74766.1; -.
KW Complete proteome.
SQ SEQUENCE 573 AA; 63889 MW; BEB5381E34A95DEF CRC64;

Alignment Scores:
Pred. No.: 138 Length: 573
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0
DB: 16 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x Q7U0D7 (1-573)

OY 59 AGAGCCAGCAAGCTCAGTAACA 36
DB 457 ArgAlaThrAsnLeuSerValThr 464

RESULT 35

Q9P8M8 PRELIMINARY; PRT; 605 AA.

AC Q9P8M8; 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE Glucose oxidase.
GN GO2.
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5061;
RN [1]_TaxID=5061;
RP SEQUENCE FROM N.A.
RC STRAIN=ACMO4;
RA Kim M., Kwon T., Yang M., Kim D.,
RT "Cloning and identification of a unique allele of the glucose oxidase
RT from a Korean wild type strain of Aspergillus niger.",

RP SEQUENCE FROM N.A.
RC TISSUE=Colon;
RA Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases
RL Strusberg R.;
DR EMBL; BC053877; AAH53877.1; -
KW Hypothetical protein.
SQ SEQUENCE 605 AA; 68046 MW; 1182534F9556DAD CRC64;

DT 01-MAY-1999 (TEMBBLrel. 10, Created)
DT 01-MAY-1999 (TEMBBLrel. 10, last sequence update)
DT 01-JUN-2003 (TEMBBLrel. 24, last annotation update)
DE En/spm-like transposon protein.
GN A12G15070.

OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Buell C.V., Mason T.M., Bowman C.L., Barnstead M.E., Feldlyum T.V.,
RA Cronin L.A., Shen M., Vanden S.E., Umeyam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhagen G.P., Preuss D., Niernann W.C., White O., Eisen J.A.,
RA Salberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana".
RL Nature 402:761-768 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC005957; AAC03364.1; -.
DR PIR; F84524; F84524.
DR InterPro; IPR004242; Transposase_21.
DR Pfam; PF02992; Transposase_21; 1.
SQ SEQUENCE 672 AA; 77763 MW; 71EBB032F1CF785C CRC64;

Alignment Scores:
Pred. No.: 135 Length: 672
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0
DB: 10 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x Q9ZUK4 (1-672)
QY 1094 ACGCTGGGACGCTGGCGGATGC 1071
DB 347 ThireulglythireulglyCys 354

RESULT 39
Q8N4M1 PRELIMINARY; PRT; 677 AA.
ID Q8N4M1;
AC Q8N4M1;
DT 01-OCT-2002 (TRENBLrel. 22, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein MGCA5474 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, Lung, and Testis;
RA Strausberg R.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC03858; AAH3858.2; -.
DR InterPro; IPR007603; DUF580.
DR Pfam; PF04515; DUF580.1.
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 677 AA; 75876 MW; 115675C8CC2D02E CRC64;

Alignment Scores:
Pred. No.: 135 Length: 677
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0

Query Match: 1.24% Indels: 0
DB: 4 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x Q8N4M1 (1-677)
QY 1493 TGGCGGTGCGCTGCTGCTGTT 1470
DB 588 TrpAlaValProLeuLeuVal 595

RESULT 40
Q65201 PRELIMINARY; PRT; 692 AA.
ID Q65201;
AC Q65201;
DT 01-AUG-1998 (TRENBLrel. 07, Created)
DT 01-AUG-1998 (TRENBLrel. 07, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Acyl-CoA oxidase (EC 1.3.3.6).
GN ACX2.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia; TISSUE=seedling hypocotyl;
RX MEDLINE=20040045; PubMed=10571860;
RA Hooks M.A., Kellis F., Graham I.A.;
RT "Long-chain acyl-CoA oxidases of Arabidopsis".
RL Plant J. 20:1-13 (1999).
DR EMBL; AF057043; AAC13497.1; -.
DR PIR; T52120; T52120.
DR GO; GO:0005777; C:peroxisome; IEA.
DR GO; GO:0003997; F:acyl-CoA oxidase activity; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR GO; GO:000635; P:fatty acid beta-oxidation; IEA.
DR InterPro; IPR00655; ACOX.
DR InterPro; IPR00690; ACOX.
DR Pfam; PF01756; ACOX; 1.
DR Pfam; PF00441; Acyl-CoA_dh; 1.
KW Oxidoreductase.
SQ SEQUENCE 692 AA; 77496 MW; 59D134B34772D24 CRC64;

Alignment Scores:
Pred. No.: 135 Length: 692
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 10 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x Q65201 (1-692)
QY 596 AGCTGACTTATCCCGAGCTC 619
DB 34 SerLeuThrLeuSerProSerLeu 41

RESULT 41
Q8ZK16 PRELIMINARY; PRT; 694 AA.
ID Q8ZK16;
AC Q8ZK16;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)
DE Putative ATP-dependent Lon protease.
GN STM4491.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=LT2 / SCSCL412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Flores L., Miller W., Stoneking T., Nian M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of *Salmonella enterica* serovar Typhimurium
LT2.";
RL Nature 413:852-856(2001).
DR EMBL: AEO08911; AL23309.1; -;
DR GO: GO:0008233; F:Peptidase activity; IEA.
KW Protease; Hypothetical protein; Complete proteome.
SQ SEQUENCE 694 AA; 77463 MW; 53BE7B95BF8B878 CRC64;

Alignment Scores:
Pred. No.: 135 Length: 694
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 16 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x Q8ZK16 (1-694)

QY 970 CCTGTCCCTGCTGCTCAGAGGT 993
DB 8 ProValProAlaValSerGluGly 15

RESULT 42
Q8U1R9 PRELIMINARY; PRT; 707 AA.
ID Q8U1R9
AC Q8U1R9;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical protein PF1136.
GN PF1136.
OS *Pyrococcus furiosus*.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC *Pyrococcus*.
OX NCBI_TaxID=2261;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Vc1 / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the *Pyrococcus furiosus* genome.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AEO10223; AL81260.1; -;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 707 AA; 80090 MW; 012ACBC9C987A6D7 CRC64;

Alignment Scores:
Pred. No.: 134 Length: 707
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 17 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x Q8U1R9 (1-707)

QY 67 AGTGAACGGAAGAGTCTGCTA 90
DB 53 SerGlyThrGluYValLeu 60

RESULT 43
P87583 PRELIMINARY; PRT; 711 AA.
ID P87583
AC P87583;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE Non-structural protein 1.
GN NS1.
OS Chikungunya virus.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Erythrovirus.
OX NCBI_TaxID=56820;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99119498; PubMed=9918883;
RA Yoo B.C., Lee D.H., Park S.M., Park J.W., Kim C.Y., Lee H.S.,
RA Seo J.S., Park K.J., Ryu W.S.;
RT "A novel parvovirus isolated from Manchurian chipmunks.";
RL Virology 253:250-258(1999).
DR EMBL: U86868; AA882733.1; -;
DR InterPro: IPR001257; Parvo_NSI.
SQ SEQUENCE 711 AA; 78578 MW; 75E64EB8563BAEB CRC64;

Alignment Scores:
Pred. No.: 134 Length: 711
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0
DB: 12 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x P87583 (1-711)

QY 1647 CTGGGGGGCCAGATATCCAGTG 1624
DB 404 LeuGlyGlyInAspIleArgVal 411

RESULT 44
Q8Y2P4 PRELIMINARY; PRT; 727 AA.
ID Q8Y2P4
AC Q8Y2P4;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein A114378.
GN A114378.
OS *Anabaena* sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_TaxID=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneo T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Matanabe A., Iriyuchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsumoto M., Muraki A.,
RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabeta S.;
RT "Complete genomic sequence of the filamentous nitrogen-fixing
cyanobacterium *Anabaena* sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
DR EMBL: AP003596; BAB76077.1; -;
DR FIRM; AB2353; AB2353.
DR InterPro: IPR008941; TPR-1like.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 727 AA; 81109 MW; 78E7D9B9FA603183 CRC64;

Alignment Scores:
Pred. No.: 134 Length: 727
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 16 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x Q8Y2P4 (1-727)

QY 1081 AGCGTGCCGCGTCCGAGCGTG 1104
DB 116 SerValProSerValProSerVal 123

```
RESULT 45
Q98S05
ID 098S05 PRELIMINARY; PRT; 752 AA.
AC 098S05;
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, last sequence update)
DE Cell division cycle protein 48 homolog.
GN CDC48.
OS Giardia theta (Cryptomonas phl.).
OC Nucleomorph.
OC Eukaryota; Cryptophyta; Cryptomonadaceae; Guillardia.
OX NCBI_TaxID=55529;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21223349; PubMed=11323671;
RA Douglas S., Zauner S., Fraunholz M., Beaton M., Penny S., Deng L.T.,
RA Wu X., Reith M., Cavalier-Smith T., Møller U.G.;
RT "The highly reduced genome of an enslaved algal nucleus.";
RL Nature 410:1091-1096(2001).
DR EMBL; AF083031; AK39773.1; -.
DR PIR; H90135; H90135.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0016787; F:Hydrolase activity; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0008151; P:cell growth and/or maintenance; IEA.
DR InterPro; IPR003959; AAA_ATPase.
DR InterPro; IPR005939; AAA_ATPase_centr.
DR InterPro; IPR003960; AAA_CDC48.
DR InterPro; IPR009010; Asp_decarb_fold.
DR InterPro; IPR00338; ATPaseVAT_N.
DR Pfam; PF00004; AAA; 2.
DR SMART; SM00382; CDC48_N; 1.
DR TIGRPFAMs; TIGR01243; CDC48; 1.
DR PROSITE; PS00674; AAA; 2.
KM ATP-binding.
SQ SEQUENCE 752 AA; 84454 MW; FE59364D5A82BD71 CRC64;

Alignment Scores:
Pred. No.: 133
Score: 8.00 Length: 752
Percent Similarity: 100.00% Matches: 8
Best Local Similarity: 100.00% Conservative: 0
Query Match: 1.24% Mismatches: 0
Indels: 0
Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x Q98S05 (1-752)
QY 771 AGCGATGCTGCGGAGCGGAGAT 748
DB 570 SeraspGlyGlyGlyalaglyasp 577

RESULT 46
Q9KIB0
ID 09KIB0 PRELIMINARY; PRT; 757 AA.
AC 09KIB0;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DE Hypothetical tonB-linked outer membrane receptor Pg13.
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=WS0.
RA Ross B., Barr I., Patterson M., Agius C., Rothel L., Margetts M.,
RA Hocking D., Webb E.;
RT "P. gingivalis polypeptides and nucleic acids.";
```

```
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=WS0;
RA Ross B.C.;
RT "Genomic analysis of Porphyromonas gingivalis for vaccine discovery.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE (BY SIMILARITY).
DR EMBL; AF237558; AAF81416.1; -.
DR GO; GO:0019867; C:outer membrane; IEA.
DR GO; GO:004872; F:receptor activity; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR008969; Carboxypeptid_reg.
DR InterPro; IPR000531; TonB_box.
DR Pfam; PF00593; TonB_dep_Rec; 1.
KM Membrane: Outer membrane: Receptor; TonB box.
SQ SEQUENCE 757 AA; 84517 MW; 224B6D5264F9D62 CRC64;

Alignment Scores:
Pred. No.: 133
Score: 8.00 Length: 757
Percent Similarity: 100.00% Matches: 8
Best Local Similarity: 100.00% Conservative: 0
Query Match: 1.24% Mismatches: 0
Indels: 0
Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x Q9KIB0 (1-757)
QY 1187 TTCTGCTCTTCAGCGCTGTCGCG 1164
DB 16 LeuLeuSerSerSerLeuSerAla 23

RESULT 47
Q91C22
ID 091C22 PRELIMINARY; PRT; 759 AA.
AC 091C22;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, last annotation update)
DE ORF1.
OS TT virus.
OC Viruses; ssDNA viruses; Circoviridae; Anellovirus.
OX NCBI_TaxID=68887;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21488921; PubMed=11601907;
RA Okamoto H., Nishizawa T., Takahashi M., Asabe S., Tsuda F.,
RA Yoshikawa A.;
RT "Heterogeneous distribution of TT virus of distinct genotypes in
RT multiple tissues from infected humans.";
RL Virology 288:358-368(2001).
DR EMBL; AB060595; BAB69912.1; -.
DR InterPro; IPR004219; TTVirus_Unk.
DR Pfam; PF02956; TT_ORF1; 1.
SQ SEQUENCE 759 AA; 89860 MW; 5B79DFD71A37010D CRC64;

Alignment Scores:
Pred. No.: 133
Score: 8.00 Length: 759
Percent Similarity: 100.00% Matches: 8
Best Local Similarity: 100.00% Conservative: 0
Query Match: 1.20% Mismatches: 0
Indels: 0
Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x Q91C22 (1-759)
QY 389 TCTCTGTTGTTTCCTGCGCTCT 412
DB 249 SerLeuValValSerLeuAlaSer 256

RESULT 48
Q813X5
```

ID 08L3X5 PRELIMINARY; PRT; 773 AA.
 AC 08L3X5;
 DT 01-OCT-2002 (TEMBLrel. 22, Created)
 DT 01-OCT-2002 (TEMBLrel. 22, Last sequence update)
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
 DE Putative reverse transcriptase.
 GN OSJNA0036D19.11 OR OSJNBA0050E08.4.
 OS Oryza sativa (Rice), and
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzoideae; Oryzae;
 OC NCBI_TaxID=4530, 39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=O.sativa, and O.sativa (japonica cultivar-group);
 RA Wang R.A., Yu Y., Yang T.J., Nah G., Soderlund C., Chen M., Kim H.-R.,
 RA Rambo T., Sakai C., Henry D., Oates R., Simmons J.;
 RT "Rice Genomic Sequence."
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=O.sativa (japonica cultivar-group); STRAIN=cv. Nipponbare;
 RA The Rice Chromosome 10 Sequencing Consortium;
 RT "In-depth view of structure, activity, and evolution of rice
 chromosome 10."
 RL Science 300:1566-1569(2003).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=O.sativa (japonica cultivar-group); STRAIN=cv. Nipponbare;
 RA Buehl C.R., Ming R.A., McCombie W.R., Messing J., Yuan Q.;
 RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC116600; AA004143.1; -
 DR EMBL; AC098696; AA004203.1; -
 DR EMBL; AB017059; AAP52227.1; -
 DR Gramene; Q8L3X5; -
 DR GO; GO:0003964; F:RNA-directed DNA polymerase activity; IEA.
 KM RNA-directed DNA polymerase
 SO SEQUENCE 773 AA; 85310 MM; DBSDCC866CC086A3 CRC64;

Alignment Scores:
 Pred. No.: 133 Length: 773
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.24% Indels: 0
 DB: 10 Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x Q8L3X5 (1-773)
 QY 1957 TGGCCACTTTCCTCCAGCTCTTCCC 1934
 DB 303 CyaproleuserProAlaLeupro 310

RESULT 49
 O8UGX6 PRELIMINARY; PRT; 789 AA.
 ID O8UGX6;
 AC O8UGX6;
 DT 01-JUN-2002 (TEMBLrel. 21, Created)
 DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
 DE Na+/H+ antiporter.
 GN NMHA OR ATU0909 OR AGR_C_1658.
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
 OC NCBI_TaxID=176299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=21608550; PubMed=11743193;
 RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boyce D. Sr.,

RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
 RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
 RA Raymond C., Rouse G., Saenphitachak C., Wu Z., Romero P., Gordon D.,
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung W., Krespan W., Perry M.,
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
 RA Neiser E.W.;
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
 C58."
 RL Science 294:2317-2323(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=21608551; PubMed=11743194;
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
 RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
 RA Houmlet K., Gordon J., Vaudin M., Iarchenko O., Epp A., Liu F.,
 RA Mollam C., Allinger M., Doughy D., Scott C., Lappae C., Markelz B.,
 RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,
 RA Cielo C., Slater S.;
 RT "Genome sequence of the plant pathogen and biotechnology agent
 Agrobacterium tumefaciens C58."
 RL Science 294:2323-2328(2001).
 DR EMBL; AE009056; AA141923.1; -
 DR EMBL; AE008022; AA86713.1; -
 DR PIR; AE2688; AE2688.
 DR PIR; H97469; H97469.
 DR GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.
 DR GO; GO:0006120; P:mitochondrial electron transport, NADH to u. . .; IEA.
 DR InterPro; IPR003918; NADH oxidred4.
 DR InterPro; IPR003916; NADH oxidred5.
 DR InterPro; IPR001750; Oxidored_q1.
 DR InterPro; IPR001516; Oxidored_q1_N.
 DR Pfam; PF00361; Oxidored_q1; 1.
 DR Pfam; PF00662; Oxidored_q1_N; 1.
 DR PRINTS; PR01434; NADHGNAS5.
 DR PRINTS; PR01437; NUOXDRPAS4.
 KM Complete proteome.
 SO SEQUENCE 789 AA; 84323 MM; 0329652251P2DSFB CRC64;

Alignment Scores:
 Pred. No.: 132 Length: 789
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.24% Indels: 0
 DB: 16 Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x O8UGX6 (1-789)
 QY 1869 CTTTCAAGGGGCTCTTCATG 1846
 DB 340 LeuphelysGlyAlaLeuPhewet 347

RESULT 50
 O92RA3 PRELIMINARY; PRT; 791 AA.
 ID O92RA3;
 AC O92RA3;
 DT 01-DEC-2001 (TEMBLrel. 19, Created)
 DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
 DE Putative cation efflux system protein.
 GN PHA2 OR R00992 OR SMC00051.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Sinorhizobium/Ensifer group; Sinorhizobium.
 OC NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RA MEDLINE=21396507; PubMed=11481430;
 RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
 RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
 RA Godrie T., Goffeau A., Kahn D., Kies E., Lelaure V., Maury D.,

RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ramsperger U.,
 RA Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.,
 RT "Analysis of the chromosome sequence of the legume symbiont
 RT *Sinorhizobium meliloti* strain 1021.",
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
 DR EMBL; AL591785; CAC45564.1; -.
 DR GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.
 DR GO; GO:0006120; P:mitochondrial electron transport, NADH to u. . .; IEA.
 DR InterPro; IPR003918; NADhub_oxred4.
 DR InterPro; IPR003916; NADhub_oxreds.
 DR InterPro; IPR001750; Oxidored_q1.
 DR InterPro; IPR001516; Oxidored_q1.
 DR Pfam; PF00361; Oxidored_q1; 1.
 DR Pfam; PF00662; Oxidored_q1_N; 1.
 DR PRINTS; PR01434; NADHDGNASE5.
 DR PRINTS; PR01437; NUOXDRDTASE4.
 KW Complete proteome.
 SQ SEQUENCE 791 AA; 84181 MW; 7D833BD253BD9FF3 CRC64;

Alignment Scores:

Pred. NO.:	132	Length:	791
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.24%	Indels:	0
DB:	16	Gaps:	0

US-10-029-345A-108_COPY_532_2532 (1-2001) x Q92RA3 (1-791)

QY 1869 CTTTCAAGGGGCTCTTCATG 1846

DB 338 LeuphelysGlyAlaLeuphemeT 345

Search completed: June 21, 2004, 13:13:30
 Job time : 148.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 21, 2004, 12:46:21 ; Search time 94.5 Seconds
(without alignments)
11965.663 Million cell updates/sec

Title: US-10-029-345A-108_COPY_532_2532

Perfect score: 667

Sequence: 1 gttgcatcgtgcccacatgagat.....tggaaatcattgaggtctcc 2001

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Word size: 1

Total number of hits satisfying chosen parameters: 324566

Minimum DB seq length: 473

Maximum DB seq length: 200000000

Post-processing: listing first 75 summaries

Command line parameters:

-MODE=frame+np.model -DEV=xlp
-Q=/cgn2.1/USPTO.spool_p/US10029345/runat_21062004_124614_5672/app_query.fasta_1.2183
-DB=A_Geneseq_29345a04 -QFMT=faatan -SUFFIX=oligo.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo.rag -TRANS=human40.cdi
-LIST=75 -DOCALLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=50 -MODE=LOCAL
-OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=473 -MAXLEN=200000000
-USER=US10029345_@CGN_1.1.81@runat_21062004_124614_5672 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

A_Geneseq_29345a04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	665	99.7	665	ABR52381	ABR52381 Protein r
2	665	99.7	665	ABR52407	ABR52407 Protein r
3	472	70.8	517	AAU79159	AAU79159 Human dua
4	472	70.8	665	AAE04834	AAE04834 Human SGP
5	472	70.8	665	AAU09016	AAU09016 Human dua
6	472	70.8	665	AAU79156	AAU79156 Human dua
7	472	70.8	665	AAU09946	AAU09946 Protein s
8	472	70.8	665	AAU75789	AAU75789 Human pro
9	472	70.8	665	ABR97946	ABR97946 Human pro
10	472	70.8	665	AAU79929	AAU79929 Human dua

11	472	70.8	665	5	ABR97291	ABR97291 Novel hum
12	472	70.8	665	5	ADA54744	ADA54744 Human pro
13	452	67.8	665	5	AAU79161	AAU79161 Human dua
14	421	63.1	665	5	AAU79162	AAU79162 Human dua
15	350	52.5	665	5	ABR52352	ABR52352 Protein r
16	318	47.7	666	4	AAAB0335	AAAB0335 Human pro
17	281	42.1	664	4	ABR52424	ABR52424 Protein r
18	258	38.7	672	4	AAU52744	AAU52744 Human pro
19	64	9.6	660	5	ABR52385	ABR52385 Protein r
20	20	3.0	679	5	ABE08458	ABE08458 Novel pro
21	20	3.0	625	5	ABR52382	ABR52382 Protein r
22	20	3.0	625	5	ABR52350	ABR52350 Protein r
23	20	3.0	625	6	ABG73440	ABG73440 Human dua
24	20	3.0	663	2	AAW29150	AAW29150 Dual-spec
25	20	3.0	663	5	ABR52351	ABR52351 Protein r
26	13	1.9	476	4	ABR63527	ABR63527 Drosophila
27	12	1.8	436	4	ABG00724	ABG00724 Novel hum
28	10	1.6	498	6	ABU29344	ABU29344 Protein e
29	8	1.2	473	7	ADA44807	ADA44807 CD4/TCR C
30	8	1.2	474	3	AAU59170	AAU59170 CD4-Ig fu
31	8	1.2	481	1	AAU93011	AAU93011 Genetic C
32	8	1.2	481	3	AAU19510	AAU19510 CD4-Igm E
33	8	1.2	481	3	AAU51081	AAU51081 Human fus
34	8	1.2	481	3	AAU59171	AAU59171 CD4-Ig fu
35	8	1.2	502	7	ADG08675	ADG08675 Novel pro
36	8	1.2	509	2	AAU00158	AAU00158 eCD4-SCFV
37	8	1.2	509	2	AAU26792	AAU26792 Mouse TIE
38	8	1.2	509	2	AAU7553	AAU7553 Amino ac1
39	8	1.2	509	2	AAU23734	AAU23734 TIE ligand
40	8	1.2	509	3	AAU90398	AAU90398 Mouse TIE
41	8	1.2	509	7	ADD69365	ADD69365 Human Ang
42	8	1.2	519	7	AAU20152	AAU20152 Human CD4
43	8	1.2	524	1	AAU94703	AAU94703 Sequence
44	8	1.2	525	3	AAU51289	AAU51289 Arabidops
45	8	1.2	525	3	AAU22387	AAU22387 Arabidops
46	8	1.2	530	2	AAU26783	AAU26783 CD4-IgG2
47	8	1.2	530	2	AAU46679	AAU46679 CD4-IgG2
48	8	1.2	530	3	AAU85080	AAU85080 CD4-IgG2
49	8	1.2	530	4	AAU67323	AAU67323 CD4-IgG2
50	8	1.2	530	4	AAU80884	AAU80884 Human CD4
51	8	1.2	530	6	ABG1123	ABG1123 CD4-Immun
52	8	1.2	532	2	AAU27278	AAU27278 CD4-gamma
53	8	1.2	532	2	AAU78678	AAU78678 T-cell re
54	8	1.2	532	2	AAU89458	AAU89458 CD4:eta f
55	8	1.2	532	2	AAU02215	AAU02215 CD4:T-cel
56	8	1.2	532	2	AAU83141	AAU83141 Chimeric
57	8	1.2	534	2	AAU26531	AAU26531 Sequence
58	8	1.2	534	5	AAU71495	AAU71495 P. amygd
59	8	1.2	540	2	AAU70460	AAU70460 South Afr
60	8	1.2	540	2	AAU70462	AAU70462 Girwood
61	8	1.2	540	2	AAU70464	AAU70464 Sindbis v
62	8	1.2	549	2	AAU04920	AAU04920 Immunopro
63	8	1.2	557	2	AAU04923	AAU04923 Immunopro
64	8	1.2	574	2	AAU04919	AAU04919 Immunopro
65	8	1.2	575	2	AAU27276	AAU27276 CD4:zeta
66	8	1.2	575	2	AAU78676	AAU78676 T-cell re
67	8	1.2	575	2	AAU89456	AAU89456 CD4:zeta
68	8	1.2	575	2	AAU02213	AAU02213 CD4:T-cel
69	8	1.2	575	2	AAU83140	AAU83140 Chimeric
70	8	1.2	576	5	AAU21625	AAU21625 Human gen
71	8	1.2	577	2	AAU04924	AAU04924 Immunopro
72	8	1.2	584	4	AAU89458	AAU89458 Novel hum
73	8	1.2	585	4	AAU89457	AAU89457 Novel hum
74	8	1.2	590	6	ABU07697	ABU07697 Viral coa
75	8	1.2	605	2	AAU04724	AAU04724 Glucose o

ALIGNMENTS

RESULT 1
ABR52381
ID ABR52381 standard; protein; 665 AA.
XX

AC ABR52381;
 DT 19-JUN-2003 (first entry)
 XX
 DE Protein relating to the invention SEQ ID NO: 109.
 XX
 KW antiproliferative; hepatotropic; nephrotropic; antiarthritic;
 KW antiproliferative; cardiant; cytostatic; gene therapy; liver disease;
 KW proliferative disorder; renal failure; cardiovascular disorder;
 KW immunological disorder; arthritis; psoriasis; congenital heart defect;
 KW congestive heart failure; cancer; human; mouse; yeast; rat; fruitfly.
 OS Homo sapiens.
 XX
 PN WO200257460-A2.
 XX
 PD 25-JUL-2002.
 XX
 PF 20-DEC-2001; 2001MO-US050459.
 XX
 PR 20-DEC-2000; 2000US-0256668P.
 PR 30-MAR-2001; 2001US-0280186P.
 PR 01-MAY-2001; 2001US-0287735P.
 PR 05-JUN-2001; 2001US-0295848P.
 PR 25-JUN-2001; 2001US-0300465P.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Jackson DG, Feder J, Nelson T, Mintier G, Ramanathan C, Lee L,
 PI Stiemers N, Bol D, Schieven G, Finger J, Todderud CG, Bassolino D,
 PI Krystek S, Mcatee P, Suchard S, Banas D,
 XX
 DR WPI: 2002-599721/64.
 DR N-PSDB; ACC60559.
 XX
 PT Novel polynucleotides encoding human phosphatase polypeptides useful in
 PT the prevention or treatment of e.g. proliferative and cardiovascular
 PT disorders.
 XX
 PS Claim 5; Fig 12; 801pp; English.
 XX
 CC The invention relates to a novel isolated nucleic acid comprising a
 CC polynucleotide having a nucleotide sequence selected from 40
 CC polynucleotides fully defined in the specification. The polynucleotide of
 CC the invention has antiproliferative, hepatotropic, nephrotropic,
 CC antiarthritic, antipsoriatic, cardiant, and cytostatic activity. The
 CC polynucleotide may have a use in gene therapy. A polynucleotide or
 CC polypeptide of the invention is useful for preventing, treating or
 CC ameliorating a medical condition, e.g. a proliferative disorder. They are
 CC also useful for treating e.g. liver disease, renal failure, immunological
 CC disorders including arthritis and psoriasis, cardiovascular disorders
 CC such as congenital heart defects and congestive heart failure, and
 CC cancer. A method of the invention is useful for diagnosing a pathological
 CC condition or susceptibility to a condition in a subject. The present
 CC sequence is used in the exemplification of the invention
 XX
 SQ Sequence 665 AA;
 XX
 Alignment Scores:
 Pred. No.: 0 Length: 665
 Score: 665.00 Matches: 665
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 99.70% Indels: 0
 DB: 5 Gaps: 0
 US-10-029-345A-108_COPY_532_2532 (1-2001) x ABR52381 (1-665)
 QY 7 ATGGCCCATGAGTGAATGGAATCTCAATTTGTAAGAGAGTGTGGTCTGCTGGAA 66
 Db 1 MetAlHisIeGIuMetIleGIyThrGlnIleValIThrGluArgLeuValAlaIleuIeuIu 20
 QY 67 AGTGAACGGAATAAGTCTGCTAAATTGATAGCCGCGCAATTTGTGAATACATACATCC 126

Db 21 SerGIyThrGIuysValIleuLeuIleAspSerArgProPheValGIuIyrAsnThrSer 40
 QY 127 CACATTTTGAAGCCATTTATATCACTGCTCCAGCTTATAGAGCAAGTTGCAACG 186
 Db 41 HisIleuGIuAlaIleAsnIleAsnIleAsnIleAsnIleAsnIleAsnIleAsnIle 60
 QY 187 GACAAAGTGTAAATTACAGAGCTCATCCAGCAATTCAGCAAACTATAGGTTACAT 246
 Db 61 AspIysValIleuIleThrGIuIleuIleGlnIleAsnIleAsnIleAsnIleAsnIle 80
 QY 247 TGCAGTCAGAAAGTTGATTTACATCAAGCTCCCAAGATGTTCCCTCTCTTCA 306
 Db 81 CysSerGIuIysValIleuValIleuValIleuValIleuValIleuValIleuValIleu 100
 QY 307 GACTGTTTCTACCTGTAATCTGAGTAACTGAGAAAGAGCTTCAATCTGTTCACTG 366
 Db 101 AspCysPheIleuThrValIleuIleuGIuIysSerPheAsnSerValIleuIleu 120
 QY 367 CTTCAGGTGGGTTTGTGAGTCTCTGCTGTTTCCCTGGCTCTGAGAGAAATCC 426
 Db 121 LeuAlaGIyGIyPheAlaGIuPheSerArgCysPheProGIyLeuGIyGIyGIySer 140
 QY 427 ACTAGTCCCTACCTGATTTCTCAGGCTTCTTACCTGTTGCAACATTTGGCCCAAC 486
 Db 141 ThrIleuValProIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 160
 QY 487 CCAATTTTCCCAATCTTATCTTGGCTGCGCAGGAGATGCTCCCAACAGAGCTGATA 546
 Db 161 ArgIleuProIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 180
 QY 547 CAGCAGAAATGGAGTTGATGTTAAATGCTTAACTGTTCCAAAGCTGACTTT 606
 Db 181 GlnGIuIleuGIyIleGIyValIleuAsnIleSerIleuIleuIleuIleuIleuIleu 200
 QY 607 ATCCCGAGTCTCAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 666
 Db 201 IleProGIuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 220
 QY 667 CCGTGTGGGCAAAATGATGATTTGATGAGAAAGCAAAAGCTTCAATGATGTT 726
 Db 221 ProIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 240
 QY 727 CTATGCACTGTGTTAGGAGTCTCCGCTCCGCAATGCTTATGCTTATGCTTATGCT 786
 Db 241 LeuValHisCysIleuAlaGIyIleSerArgSerAlaIleIleAlaIleAlaIleIle 260
 QY 787 AAGGATGACATGTTTATGATGAGCTTACAGATTTGTGAAAGAAAGAAAGCTTACT 846
 Db 261 LysArgMetAspMetSerIleuAspGIuAlaIleuIleuIleuIleuIleuIleuIleu 280
 QY 847 ATATCTGCAAACTTCAATTTTCTGGGCAATCTTGGATTTAGAGAAAGATTAAGAC 906
 Db 281 IleSerProAsnIleAsnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 300
 QY 907 CAGACTGAGCATCAGGCGCAAAAGAGCAAACTCAAGCTGCTGCTGAGAGAGCAAT 966
 Db 301 GlnThrGIyAlaSerGIyProIysSerIyIleuIleuIleuIleuIleuIleuIleuIleu 320
 QY 967 GAACCTGTCTGCTGCTTCTCAAGAGGTGACAGAAAGAGAGAGAGAGAGAGAGAGAG 1026
 Db 321 GluProValProAlaValSerGIuIleuIleuIleuIleuIleuIleuIleuIleuIleu 340
 QY 1027 TGTGCGACTGTGCTTACCTCAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1086
 Db 341 CysAlaIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 360
 QY 1087 CCAAGCTGCCAGGCTGAGCGCTGCTGTTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1146
 Db 361 ProSerValProSerValGIuProSerIleuIleuIleuIleuIleuIleuIleuIleu 380
 QY 1147 AGTGGCTGCACTGTTCCGAG 1206

QY 247 TGCAGTCAAGAGTTGATGTTTACGATCAAAAGCTCCCAAGATGTTGCTCTCTCTTCA 306
Db 81 CysSerGlnIysValValValIyrrAspGlnSerSerGlnSepValAlaSerLeuSer 100
QY 307 GACTGTTTCTCACTGACTTCTGGGTAACCTGGAAGAAGCTTCAACTCTGTACACTG 366
Db 101 AspCysHeuThrValLeuGlnGlyyblEuGlnLysSerPheksnSerValHibEu 120
QY 367 CTTCAGAGTGGGTTTGCTGAGTTCTGCTGGTTTCCCGGCTCGTGAAGAGAAATCC 426
Db 121 LeuAlaGlyIyhealagIuPheSerAlaGlyCysGlnAAspValLeuAsnLysGlnLysSer 140
QY 427 ACTCTGCTCCCTACCTGCAATTTCTGACCTGCTTACCTGTTGCCAAGATTGGGCAAC 486
Db 141 ThrLeuValProThrCysIleSerGlnProCysLeuProValAlaAsnIleGlyProThr 160
QY 487 CGAATCTTCCCAATTTTATCTTGTGGCCGCAAGATGTCCTCAACAAGAGCTGTA 546
Db 161 ArgIleLeuProAsnLeuIyrrLeuGlyCysGlnAAspValLeuAsnLysGlnLysSer 180
QY 547 CAGCAGATGGGATTTGTTATGTTAAATGCACTATACCTGTCGCAAGCTGACTT 606
Db 181 GlnGlnAsnGlyIleGlyValValLeuAsnIleSerIyrrThrCysProLysProAspPhe 200
QY 607 ATCCCGAGTTCATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 666
Db 201 IleProGlnSerHisPheLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaIyrrIleMet 220
QY 667 CCGTGGTGGACAAATCAGATGATTTCTGAGAAGAAAGCCCTCAATGATGTT 726
Db 221 ProThrLeuAspLysSerValAspPheIleGlyValAlaLysAlaSerAsnGlyCysVal 240
QY 727 CTAGTGCAGTGTGTTAGCTGGGATCTCCGCTCCGCCACATCGCTATCGCTTACATCATG 786
Db 241 LeuValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaIyrrIleMet 260
QY 787 AAGAGATGACATGCTTTAGATGAGCTTACAGATTTGTAAGAAAGAAAGAAAGCTTACT 846
Db 261 LysArgMetAspMetSerLeuAspGlnAlaIyrrArgPheValLysGlnLysArgProThr 280
QY 847 ATATCTCCAACTTCAATTTTCTGGGCCAACTCTGACTATGAGAAGAAAGCTTAAAG 906
Db 281 IleSerProAsnPheAsnPheLeuGlyGlnLeuLeuAspIyrrGlnLysIleLysAsn 300
QY 907 CAGACTGAGCATCAGGAGGCAAGAGAACTCAACTGTGTGCACTGGAAGAGCCAAAT 966
Db 301 GlnThrGlyAlaSerGlyProLysSerLysLeuLysLeuLeuHibLeuGlnLysProAsn 320
QY 967 GAACCTGTCTCTGCTTCTCAGAGGCTGACAGAAAGAGAGAGCCCTCAGTCCACCC 1026
Db 321 GluProValProAlaValSerGlnGlyGlnLysSerGlnThrProLeuSerProPro 340
QY 1027 TGTGCCAGCTCTGCTCAAGAGGAGAGAGCAAGAAAGCCGTGCATCCGCGCAGCTG 1086
Db 341 CysAlaAspSerAlaThrSerGlnAlaAlaGlyGlnArgProValHibProAlaSerVal 360
QY 1087 CCCAGCGTGGCCAGCGGTGAGCGCTGCTGTAGAGAGAGAGCCGCTGATCAAGCGCTG 1146
Db 361 ProSerValProSerValGlnProSerLeuLeuGlnAspSerProLeuValGlnAlaLeu 380
QY 1147 AGTGGCTGCACTGTCCCGCAGACAGGCTGGAAGACAGCAATAGCTTCAAGCTTCTTC 1206
Db 381 SerGlyLeuHibLeuSerAlaAspArgLeuGlnAspSerAsnLysLeuLysArgSerPhe 400
QY 1207 TCTTGATATCAAAATCAGTTTCATATTCAGCGAGAGAGAGATCCTTACATGCTTC 1266
Db 401 SerLeuAspIleLysSerValSerIyrrSerAlaSerMetAlaAlaSerLeuHibGlyPhe 420
QY 1267 TCCCTATCAGAAAGATGCTTTGGAATATCAAACTTCCACTTCTGAGTGGAGCCAC 1326
Db 421 SerSerSerGlnAspAlaLeuGlnIyrrIyrrLysProSerThrThrLeuAspGlyThrAsn 440
QY 1327 AAGCTATGCCAGTTCTCCCTGTTTCAAGAACTATTCGAGAGAGATCCCGAAACCAAGTCTT 1386

Db 441 LysLeuCysGlnPheSerProValGlnGlnLysSerGlnGlnThrProGlnThrSerPro 460
QY 1387 GATTAAGAGGAAGCCAGCATCCCGAAAGAGCTGACAGCCGAGCCTTACAGACCCAG 1446
Db 461 AspLysGlnGlnLysLeuSerIleProLysLysLeuGlnThrAlaArgProSerAspSerGln 480
QY 1447 AGCAAGCATTTGATTCGCTGCAAGACAGACAGCTGGCAACCGCCAGAGCTCTTTTA 1506
Db 481 SerLysArgLeuHibSerValArgHisSerSerSerGlyThrAlaGlnArgSerLeuLeu 500
QY 1507 TCTCCACTGCATCGAAGTGGAGCGGTGAGAGCAATTAACAACAGCTTCTTTGGGC 1566
Db 501 SerProLeuHibArgSerGlySerValGlnAspAsnIyrrHisThrSerPheLeuPheGly 520
QY 1567 CTTCACAGGACAGACAGACCTCAAGAGTCTGTGAGCTGGAGCTTAAAGGCTGGCAC 1626
Db 521 LeuSerThrSerGlnGlnHibLeuThrLysSerAlaGlyLeuGlyLysGlyIyrrHis 540
QY 1627 TCGATATCTTGGCCCCCAGACCTTACCCCTTCCCTGACAGACAGCTGGTATTTGGC 1686
Db 541 SerAspIleLeuAlaProGlnThrSerThrProSerLeuThrSerSerIyrrPheAla 560
QY 1687 ACAGAGTCTCAGACTTACTTCTGCTCAGCCATCTTACGAGAGGACAGTTCATCTCT 1746
Db 561 ThrGlnSerThrAspPheIyrrSerAlaSerAlaIleIyrrGlySerAlaSerIyrrSer 580
QY 1747 GCTTACACTTCAGACGCTGCCACTTGGGAGAGCCAAAGCTTATCTTGGGAGAGCGG 1806
Db 581 AlaIyrrSerCysSerGlnLeuProThrCysGlyAspGlnAlaIyrrSerValArgArg 600
QY 1807 CAGAAAGCAACTGACAGAGCTGACTCGGCGGAGAGCTGGCATGAGAGAGCCCTTGA 1866
Db 601 GlnLysProSerAspArgAlaAspSerArgArgSerTrpHisGlnLysSerProPheGln 620
QY 1867 AAGCAGTTTAAACGACAGAGCTGCAAAATGAAATTTGGAGAGGCTATGTCAGAGAAC 1926
Db 621 LysGlnPheLysArgHisArgSerCysGlnMetGlnPheGlyGlnSerIleMetSerGlnAsn 640
QY 1927 AGGTCAAGGAGAGAGCTGGGAGAAAGTGGCAGTCACTTACGCTTTCGGGAGCATGGA 1986
Db 641 ArgSerArgGlnGlnLeuGlyLysValGlySerGlnSerSerPheSerGlySerMetGln 660
QY 1987 ATCATTTAGAGTCTCC 2001
Db 661 IleIleGlnValSer 665

RESULT 3
AAU79159 standard; protein; 517 AA.
AAU79159;
02-JUL-2002 (first entry)

Human dual-specificity phosphatase-3 (DSP-16) alternative form protein.
Human; dual-specificity phosphatase-3; DSP-16; autoimmune disease;
malignant-activated protein kinase; MAP; Duchenne muscular dystrophy;
cancer; graft-versus-host disease; allergy; metabolic disease;
abnormal cell growth; abnormal cell proliferation; contact inhibition;
cell cycle abnormality; anchorage independent cell growth; apoptosis;
intercellular adhesion; DSP-16 modulator; chromosome 12p.

Homo sapiens.
OS
XX
FH Key Location/Qualifiers
FT Domain 94..103
XX /label= Active_site_domain
XX WO200226997-A2.
XX
XX 04-APR-2002.

XX 25-SEP-2001; 2001WO-US030124.
 PF XX
 XX
 PR 26-SEP-2000; 2000US-0235487P.
 XX
 XX
 PA (CEPT-) CEPTYR INC.
 XX
 XX
 PI Lucche RM, Wei B;
 XX
 XX
 DR WPI; 2002-315802/35.
 DR N-PSDB; ABK48378.
 XX
 XX
 PT New DSP-16 polypeptide, useful for identifying modulators of its
 PT activity, which can be used in the treatment of disorders such as
 PT Duchenne muscular dystrophy, or cancer.
 XX
 XX
 PS Claim 50; Fig 4; 87pp; English.

CC The present invention relates to a new polypeptide, DSP-16, having a 665
CC amino acid sequence, given in the specification, or a variant having at
CC least 50 % identical residues, which retains the ability to
CC dephosphorylate an activated mitogen-activated protein (MAP) kinase. The
CC invention can be used for identifying agents which modulate DSP-16
CC activity, for modulation of a proliferative response in a cell, survival
CC of a cell, or differentiation of a cell. The cell displays contact
CC inhibition of cell growth or anchorage independent growth and may display
CC altered intercellular adhesion. The agent may modulate apoptosis, or the
CC cell cycle. The identified modulators can be used to treat Duchenne
CC muscular dystrophy, cancer, graft-versus-host disease, autoimmune
CC diseases, allergies, metabolic diseases, abnormal cell growth, abnormal
CC cell proliferation, and cell cycle abnormalities. The present amino acid
CC sequence represents the human dual-specificity phosphatase-3 (DSP-16)
CC alternative form protein of the invention. This sequence is encoded by
CC the human DSP-16 gene located on chromosome 12p

SQ Sequence 517 AA;

Alignment Scores:		
Pred. No.:	0	517
Score:	472.00	Matches: 472
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	70.76%	Indels: 0
DB:	5	Gaps: 0

US-10-029-345A-108 COPY 532 2532 (1-2001) X AAU79159 (1-517)

QY 556 A C C T G T C C A A A G C C T G A C T T T A T C C C G A G N T C A T T T C T G C G G T G T C C T G T A A T A G C 645

Db 46 T h r C y s P r o L y s P r o A s p P h e I l e P r o L u s e r I l a s p h e L e u A l a G y l P r o V a l A s n A s p 65

QY 646 A G C T T T T G A G A A A T T T G C C G T G T T G C A A A T C A G T A G A T T T C A T T G A G A A A G C A 705

Db 66 S e r P h e C y s G l u L y s I l e L e u P r o I l e u P r o I l e u A m p L y s S e r V a l A s p P h e I l e G l u L y a l a 85

QY 706 A A A G C C T C C A A T G A T G T G T T C T A G T G A C A C T G T T T A G C T G G A T C T C C C G T C G G C A C C 765

Db 86 L y s A l a S e r A s n G l y C y s V a l L e u V a l I s c y s L e u A l a G l y I l e S e r A r s e r L a t h r 105

QY 766 A T C G C T A T C G C C T T A C A T C A T G A A G A G A T G G A C A T G C T T T A G A T G A A G C T T A C A G A T T 825

Db 106 I l e a l a I l e a l a r T y l l e m e t h y l a r g h e t a p m e t s e r L e u a m p G l u A l a T y r A r g h e 125

QY 826 G T G A A A G A A A A A A G C C T A C T A T A T A T C T C C A A C T T C A A T T T T C T G G G C C A A C T C C T G A C 885

Db 126 V a l l y s G l u L y s A r g P r o T h r I l e S e r P r o a n P h e a n P h e L e u G l y G l u L e u A s p 145

QY 886 T A T G A A G A A A A C A T T A A G A C C A G C T G A G A C A T A G G G C C A A A G A A C T A A G C T G 945

Db 146 T y r G l u L y s L y s I l e L y s A s n G l n t h r G l y A l a S e r G l y P r o L y s e r L y s L e u L y s L e u 165

QY 946 C T G C A C C T G A G A A G C C A A T G A A C C T G T C C C T G C T G T C T C A G A G G T G A C A G A A A G C 1005

[illegible]

RESULT 4
AAE04834
ID AAE04834 standard; protein; 665 AA.

XX AAE04834;
AC
XX 10-SEP-2001 (first entry)
DT
XX
XX Human SGP002 phosphatase polypeptide.
DE
XX
XX Human; SGP002 phosphatase polypeptide; phosphatase-related disease;
KW immune-related disorder; organ transplant rejection;
KW infection; diabetes; pain; sexual dysfunction; Alzheimer's disease;
KW metabolic disorder; hematopoietic cancer; mood disorder; cardiac;
KW Parkinson's disease; multiple sclerosis; amyotrophic lateral sclerosis;
KW cardiovascular disease; brain; neuronal-associated disease; dyskinesia;
KW attention disorder; cognition disorder; psychotic disorder; cyostatic;
KW neurological disorder; virulence; neurotropic; cerebroprotective; therapy;
KW neuroprotective; antibacterial; vulnerrary; tranquilizer; analsthetic;
KW hypotensive; immunosuppressive; antiporatic; analgesic; hypertensive;
KW antifungal; dual specificity phosphatase; DSP; MAP kinase phosphatase;
KW MKP; migraine; chromosome 12p11.1-p12.1.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH Domain 1..173
FT /label=Catalytic_domain
FT 158..297
FT Domain /label=Phosphatase_domain
XX
XX WO200146394-A2.
XX
XX 28-JUN-2001.
XX
XX 21-DEC-2000; 2000WO-US034736.
XX
XX 21-DEC-1999; 99US-017325SP.
XX 28-DEC-1999; 99US-0175766P.
XX 25-JAN-2000; 2000US-0178078P.
XX 31-JAN-2000; 2000US-0179301P.
XX
XX (SUGEN -) SUGEN INC.
XX
XX Plowman GD, Martinez R, Whyte D, Manning G, Sudarsanam S;
PI Hill RJ, Flanagan P;
XX WPI; 2001-418058/44.
XX DR N-PSDB; AAD09492.
XX
XX Novel phosphatase polypeptide useful for treating cancers, immune-related
FT diseases and disorders, cardiovascular disease, brain or neuronal-
FT associated diseases and metabolic disorders.
XX
XX Claim 7; Fig 2; 186pp; English.
XX
XX The present invention relates to phosphatase polypeptides, nucleotide
XX sequences encoding them, as well as various products and methods useful
XX for the diagnosis and treatment of various phosphatase-related diseases
XX and conditions. Substance that modulates the activity of phosphatase
XX polypeptide is used to treat immune-related diseases and disorders,
XX cardiovascular disease, brain or neuronal-associated diseases and
XX metabolic disorders, including cancers of tissues, cancers of
XX hematopoietic origin, diseases of central and peripheral nervous system,
XX Alzheimer's disease, Parkinson's disease, multiple sclerosis, amyotrophic
XX lateral sclerosis, viral infections, infections caused by prions,
XX bacteria and fungi, ocular diseases, diabetes, migraines, pain, sexual
XX dysfunction, mood disorders, attention disorders, cognition disorders,
XX hypertension, hyperextension, psychotic disorders, neurological disorders,
XX dyskinesias and organ transplant rejection. The present amino acid
XX sequence is human SGP002 phosphatase polypeptide. This sequence is
XX classified as dual specificity phosphatase (DSP) and MAP kinase
XX phosphatase (MKP). SGP002 gene maps to chromosome position 12p11.1-p12.1
XX
XX Sequence 665 AA;

Alignment Scores:

Pred. No.:	0	Length:	665
Score:	472.00	Matches:	472
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	70.76%	Indels:	0
DB:	4	Gaps:	0

US-10-029-345a-108_COPY_532_2532 (1-2001) x AAE04834 (1-665)

QY	586	ACCTGTCGCAAGCTGACTTATATCCCGAGTCTATTTCGCGTGCCTGTGAATGAC	645
DB	194	ThrCysProLysProAspPheIleProGlnSerHisPheLeuArgValProValAsnAsp	213
QY	646	AGCTTTTGAGAAATAATTTGGCGGTGGACCAATAGAGATTTCATTGAGAAAGCA	705
DB	214	SerPheCysGlnLysIleLeuProTyrLeuAspLysSerValAspPheIleGlnValAla	233
QY	706	AAAGCTCCGATGATGTTGTTCTAGTGCACTGTTAGCTGGAGATCTCCGCTCCGAC	765
DB	234	LysAlaSerAsnGlnCysValLeuValHisCysLeuAlaGlyIleSerArgSerAlaThr	253
QY	766	ATCGCTATCGCTCATCATGATGAAGATGACATGCTTTAGATGAAGCTTACAGATT	825
DB	254	IleAlaIleAlaTyrIleMetLysArgMetAspMetSerLeuAspGlnAlaTyrArgPhe	273
QY	826	GTGAAGAAAAAAGACCTACTATATCTCAAACTTCAATTTTCTGGGCGCAACTCCGAC	885
DB	274	ValLysGlnLysAspProThrIleSerProAsnPheAsnLeuGlnLeuLeuAsp	293
QY	886	TATGAGAAAGAGATTAAAGAACAGACTGAGACTCAGAGGCAAGCAACTCAAGCTG	945
DB	294	TyrGlnLysLysIleLysAsnGlnThrGlyAlaSerGlyProLysSerLysLeuLysLeu	313
QY	946	CTGCACCTGGAGAAAGCAAAATGAACCTGCTCCGCTGCTCAGAGGGGTGACAGAAAG	1005
DB	314	LeuHisLeuGlnLysProAsnGlnProValProAlaValSerGlnIleGlnLysSer	333
QY	1006	GAGAGCGCCCTCAGTCCACCTGTGCGCACTGTACTACCTCAGAGGACAGCAAGAAAG	1065
DB	334	GlnThrProLeuSerProProCysAlaAspSerAlaThrSerGlnAlaIleGlnIle	353
QY	1066	CCGTCGATCCCGCAGCGCGTCCAGCGCCAGCGTCCAGCGTCCGTTAGAGAGAC	1125
DB	354	ProValHisProAlaSerValProSerValProSerValGlnProSerLeuLeuAspSer	373
QY	1126	AGCCCGCTGTGACAGCGCTCAGTGGGCTGACCTGTCCGACAGACAGGCTGGAAAGC	1185
DB	374	SerProLeuValGlnAlaLeuSerGlyLeuHisLeuSerAlaAspArgLeuGlnAspSer	393
QY	1186	AATAGCTCAAGGTTCTCTCTCTCTGATATCAATATCAATTTTCATATTCAGCAGCATG	1245
DB	394	AsnLysLeuLysArgSerPheSerLeuAspIleLysSerValSerTyrSerAlaSerMet	413
QY	1246	GCAGCATCTTACATGAGCTTCTCTCATCAGAAAGATGGTTGGAATCTAATAACCTTCC	1305
DB	414	AlaAlaSerLeuHisGlyPheSerSerSerGlnAspAlaLeuGlnLysTyrLysProSer	433
QY	1306	ACTACTCTGGATGGAGCAACAAGCTATGCGAGTTCTCCCTGTTCGGAACCTATCGGAG	1365
DB	434	ThrThrLeuAspGlyThrAsnLysLeuCysGlnPheSerProValGlnGlnLeuSerGln	453
QY	1366	CAGACTCCGAGAAACAGGTCCTGATAGAGAGAACCCAGCATCCCAAGAGCTGCAGACC	1425
DB	454	GlnThrProGlnThrSerProAspLysGlnLysAlaSerIleProLysLysLeuGlnThr	473
QY	1426	GGCAGGCTTTAGACAGCAGCAGCAAGCATTTGCGTACGAGAACACAGCAGAGTGGC	1485
DB	474	AlaArgProSerAspSerGlnSerLysArgLeuHisSerAlaArgThrSerSerSerGly	493
QY	1486	ACCGCCGAGAGGTCCCTTTATCTTCACCTGCATCGAAAGTGGAGCGGTGAGAGCAATTAC	1545
DB	494	ThrAlaGlnArgSerLeuLeuSerProLeuHisArgSerGlySerValGlnAspAsnTyr	513

QY 1546 CACAGCAGCTCTCTTTCCGCTTTTCCACACGACGACGACGACCTCAAGAGTGTGCGC 1605
 DB 514 HstHrSerPheLeuPheGlyLeuSerThSerGlnHstLeuThrYsSerAlaGly 533
 QY 1606 CTGGAGCTTAAGAGGCTGACACTGCAATATCTTGGCCCCCAGACCTTACCTCTCCG 1665
 DB 534 LeuGlyLeuLeuPheGlyTrpHisSerAlaPheLeuAlaProGlnThrSerThrProSerLeu 553
 QY 1666 ACCAGCAGCTGTATTTTGGCCACAGAGTCTCACTTCTACTCTGCTCAGCCATCTAC 1725
 DB 554 ThSerSerTrpYrPheAlaTrnGlnSerSerHisPheYrSerAlaSerAlaIleYr 573
 QY 1726 GAGGAGGAGCCAGCTTACTGCTTACAGTGCAGCCAGCTGCGCCCTTGGAGACCA 1785
 DB 574 GlyGlySerAlaSerYrSerAlaYrSerCySerSerGlnLeuProThrCybGlyAaPglIn 593
 QY 1786 GTCTATTTCTGTGCGCAGCGGCGAGACCAAGTGAAGAGTGAAGTGAAGTGAAGTGAAG 1845
 DB 594 ValYrSerValAlaTrnAlaGlnGlnYrProSerAlaPheAlaAaPheAlaGlnSerTrp 613
 QY 1846 CATGAAGAGAGCCCTTTGAAAAGAGTTTAAAGCAGAGAGTGCAGAAATGAATTGGA 1905
 DB 614 HstGlnGlnSerProPheGlnYrGlnPheYrAlaGlnSerCybGlnMetGlnPheGly 633
 QY 1906 GAGAGCATCTATCTCAGAGAACAGGTCACGGGAAAGAGCTGGGAAAGTGGCACTGCT 1965
 DB 634 GluSerIleuSerGlnAaGlnAaGlnSerAlaGlnGlnLeuGlnYrValGlySerGlnSer 653
 QY 1966 AGCTTTTGGGCGAGCATGGAATCATTTAGAGTCTCC 2001
 DB 654 SerPheSerGlySerMetGlnIleGlnValSer 665

RESULT 5
 AAU09016
 ID AAU09016 standard; protein; 665 AA.
 XX
 AC AAU09016;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Human dual specificity phosphatase 21117.
 XX
 KW Human; dual specificity phosphatase 21117; hepatotropic; cytosolic;
 KW hematopoietic disorder; autoimmune disorder; diabetes mellitus;
 KW rheumatoid arthritis; multiple sclerosis; Crohn's disease;
 KW liver disorder; erythroid associated disorder; hemolytic anaemia;
 KW cellular proliferative; differentiative disorder; leukemia;
 KW acute myeloid leukemia; carcinoma; sarcoma; metastatic cancer;
 KW immunogen.
 XX
 OS Homo sapiens.
 XX
 FH Key
 XX
 FT Domain
 XX
 FT Domain
 XX
 FT 11..131
 FT 158..297
 FT /label= Rhodanese-like domain
 FT /label= Catalytic domain
 FT /note= "Dual specificity phosphatase catalytic domain"
 FT 242..254
 FT /label= Tyrosine-specific_protein_phosphatase_active_site
 FT
 XX
 XX WO200173059-A2.
 XX
 XX 04-OCT-2001.
 XX
 XX 23-MAR-2001; 2001WO-US009477.
 XX
 XX 24-MAR-2000; 2000US-0191858P.
 XX
 XX (MILL-) MILLENNIUM PHARM INC.
 XX
 XX Meyers RA;

XX
 DR MPI: 2001-611635/70.
 XX N-PSDB; AAS14639.
 XX
 PT New human dual specificity polypeptides and nucleic acids for diagnosis
 XX of disease and treatment of e.g. liver disorders.
 XX
 PS Claim 9; Fig 1; 143pp; English.
 XX
 CC The invention relates to two novel human dual specificity phosphatases
 CC designated 21117 and 38692, the nucleic acids encoding them (including
 CC fragments, allelic variants, their complements or nucleic acids that
 CC hybridise to them) and antibodies raised against the proteins. The
 CC antibody is useful for detecting the presence of the polypeptide, and the
 CC nucleic acid fragments are useful for detecting the presence of the
 CC nucleic acid, for diagnosing disease. Agents (e.g. the antibody and
 CC antisense sequences) are useful for modulating the activity or expression
 CC of the polypeptide/nucleic acid, useful for the treatment of e.g.
 CC hematopoietic disorders, autoimmune disorders (e.g. diabetes mellitus,
 CC rheumatoid arthritis, multiple sclerosis, Crohn's disease and many others
 CC listed in the specification) liver disorders, erythroid associated
 CC disorders (e.g. haemolytic anaemia) cellular proliferative or
 CC differentiative disorders, leukemias (e.g. acute myeloid leukaemia),
 CC carcinomas, sarcomas, metastatic cancers. Agents modulating 21117 and
 CC 38692 are also useful for modulating the proliferation, survival,
 CC migration or differentiation of a 38692 or 21117-expressing cell. The
 CC polypeptide and nucleic acids are useful for identifying modulating
 CC agents. The present sequence represents the dual specificity phosphatase
 CC 21117
 CC
 XX
 SQ Sequence 665 AA;
 Alignment Scores:
 Pred. No.: 0 Length: 665
 Score: 472.00 Matches: 472
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 70.76% Indels: 0
 DB: 4 Gaps: 0
 US-10-029-345A-108_COPY_532_2532 (1-2001) x AAU09016 (1-665)
 QY 586 ACCTGTCGAAGAGCTGATTTATCCCGAGTCTATTCCTCGTGGTGAATGAC 645
 DB 194 ThrCybProLysProAaPheHleProGlnSerHisPheLeuAlaGlnAlaProValAaNap 213
 QY 646 ACCTTTTGAAGAAAATTTTGGCGTGTGACCAATAGTATTTCACTTGAAGAAAGA 705
 DB 214 SerPheCybGlnYrIleLeuProThrPheAaPheYrSerValAaPheHleGlnYrAla 233
 QY 706 AAAGCTCCAAATGATGTGTCTAGTCACTGTTAGCTGGAGATCTCCGCTCCGACAC 765
 DB 234 LysAlaSerAaGlnGlyCybValLeuValHisCybLeuAlaGlyIleSerAlaSerAlaTrn 253
 QY 766 ATCGCTATGCTTACATCATGAAGAGATGACATGCTCTTTAGATTAACACTTACAGATT 825
 DB 254 IleAlaIleAlaTrnIleMetLysAlaMetAaPheMetSerLeuAaPheGlnAlaTrnArgPhe 273
 QY 826 GTGAAGAAAAGAAAGAAAGCTTATATCTCAAACTTCAATTTTCTGGGCAACTCTGAGAC 885
 DB 274 ValYrGlnYrAlaGlnProThrIleSerProAaPheAaPheLeuGlnLeuLeuAaP 293
 QY 886 TATGAAGAGAGATTAAGAACAGACAGTGAAGATCAGGCGCAAGAGCAAACTCAAGCTG 945
 DB 294 TyrGlnYrYrIleYrAaGlnInTrnGlyAlaSerGlyProYrSerYrLeuYrAa 313
 QY 946 CTGCACTGGAAGAACCAAAATGAACCTGCTCTGCTGTCTCAGAGGCTGACAGAAAGC 1005
 DB 314 LeuHisLeuGlnYrAlaProAaGlnProValProAlaValaSerGlnGlyIleGlnYrSer 333
 QY 1006 GAGAGCGCCCTCAGTTCACCTGTGCGCAGTCTGTACTCTCAGAGGACAGAGCAAGAG 1065
 DB 334 GluThrProLeuSerProProCybAlaAaPheSerAlaTrnSerGlnAlaIleGlnAaG 353

```
QY 1066 CCGGTGCATCCCGCAGCGTGGCCAGCGTGGCCAGCGTGGCCAGCGTGGTGTAGAGAC 1125
DB 354 ProValHisProAlaSerValProSerValGlnProSerLeuLeuGlnAsp 373
QY 1126 AGCCCCCTGTGACAGCGCGCTCAGTGGCTGCACTTCCGACAGACGCGTGGAGACAGC 1185
DB 374 SerProLeuValGlnAlaLeuSerGlyLeuHisLeuSerAlaAspArgLeuGlnAspSer 393
QY 1186 AATAAGCTACAGCGTCTCTCTCTGATCAATCAATAGTTTCAATTATTCAGCCAGCATG 1245
DB 394 AsnIlySLeuIlyArgSerPheSerLeuAspIleIlySerValSerTyrSerAlaSerMet 413
QY 1246 GACAGCATCTTTCATGAGCTTCTCTCATCAGAGAGATGCTTGGAAATACACAACTTCC 1305
DB 414 AlaAlaSerLeuHisGlyPheSerSerSerGlnAspAlaLeuGlnIlyTyrIlyProSer 433
QY 1306 ACTACTCTGGATGGAGACCAACAAGCTATGCCAGTTCTCCCTGTTCAGAACTATCGAG 1365
DB 434 ThrThrLeuAspGlyThrAsnIlyLeuGlySglnPheSerProValGlnGlnLeuSerGlu 453
QY 1366 CAGACTCCGAAACCAAGTCTGTATAGAGAGAGAGACATCCCGCAGAGCTGACACACC 1425
DB 454 GlnThrProGlnThrSerProAspIlySglnIlyAlaSerIleProIlySLeuGlnThr 473
QY 1426 GCCAGGCTTTCAGACAGCCAGACAGAGGATTCGATTCGATCAGAACCCAGCAGCAGTGC 1485
DB 474 AlaArgProSerAspSerGlnSerIlyArgLeuHisSerValArgThrSerSerSerGly 493
QY 1486 ACCGCCCGAGAGTCCCTTATATCCCACTGCATGAGAGAGCGGAGACGAGACATTCAC 1545
DB 494 ThrIlaGlnArgSerLeuLeuSerProLeuHisArgSerGlySerValGlnAspAsnIly 513
QY 1546 CACACACAGCTTCTTTTTCGCTTTCACACAGCCAGCAGACCTTCAGAACTCGCTGAC 1605
DB 514 HisThrSerPheLeuPheGlyLeuSerThrSerGlnGlnHisLeuThrIlySerAlaGly 533
QY 1606 CTGGGCGCTTAAAGGCGTGGCACTGGATATCTTGCGCCCGCAGACCTTCACCTTCCCTG 1665
DB 534 LeuGlyLeuIlySglnIlyPheHisSerAspIleLeuAlaProGlnThrSerThrProSerLeu 553
QY 1666 ACCAGCAGCTGTATTTTGGCACAGAGTCTTCACACTTACTCTGCTCAGGCACTTCAC 1725
DB 554 ThrSerSerIlyPyrPheAlaThrGlnIlySerHisPheIlySerAlaSerAlaIleIly 573
QY 1726 GAGAGCGATGCCAGTTACTCTGCTTACAGCTGCAGCCAGCTGCCACTTGGGAGACCA 1785
DB 574 GlyIlySerAlaSerIlySerAlaIlySerCysSerGlnLeuProThrCysGlyAspGln 593
QY 1786 GTCTATTCTGTGGCAGAGCGGAGAGCCAGAGTGAAGAGCTGACTCGCGGCGAGCTGG 1845
DB 594 ValIlySerValArgArgArgGlnIlySProSerAspArgAlaAspSerArgArgSerIly 613
QY 1846 CATGAAGAGAGCCCTTTGAAAAGCAATTAAACGACAGAGCTGCACAAATGGAATTTTGA 1905
DB 614 HisGlnGlnIlySerProPheGlnIlySglnPheIlySArgArgSerCysGlnMetClnIly 633
QY 1906 GAGAGCATCATGTCAGAGAAAGAGTCAACGAGAGAGCTGGGAAAGTGGCGCATGACTCT 1965
DB 634 GlnSerIleLeuSerSerIlyAsnArgSerArgGlnIlyLeuGlyValGlySerGlnSer 653
QY 1966 AGCTTTTGGGAGCATGGAATCATTTAGGTCTCC 2001
DB 654 SerPheSerGlySerMetClnIleIleGlnIlyValSer 665
RESULT 6
AAU79156
AAU79156 standard; protein; 665 AA.
XX
AC AAU79156;
XX
DT 02-JUL-2002 (first entry)
XX
```

```
DE Human dual-specificity phosphatase-3 (DSP-16) protein.
XX
KW Human; dual-specificity phosphatase-3; DSP-16; autoimmune disease;
KW mitogen-activated protein kinase; MAP; Duchenne muscular dystrophy;
KW cancer; graft-versus-host disease; allergy; metabolic disease;
KW abnormal cell growth; abnormal cell proliferation; contact inhibition;
KW cell cycle abnormality; anchorage independent cell growth; apoptosis;
KW intercellular adhesion; DSP-16 modulator; chromosome 12p.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 242..251
FT /label= Active_site_domain
PN MO200226997-A2.
PD
XX 04-APR-2002.
PP
XX 25-SEP-2001; 2001WO-0501024.
XX
XX 26-SEP-2000; 2000US-0235487P.
XX (CEPT-) CEPTYR INC.
XX
XX Luche RM, Wei B;
XX WPI; 2002-315802/35.
XX N-PSDB; ABK47596.
XX
XX New DSP-16 polypeptide, useful for identifying modulators of its
XX activity, which can be used in the treatment of disorders such as
XX Duchenne muscular dystrophy, or cancer.
XX
XX Claim 1; Fig 2; 87pp; English.
XX
XX The present invention relates to a new polypeptide, DSP-16, having a 665
XX amino acid sequence, given in the specification, or a variant having at
XX least 50 % identical residues, which retains the ability to
XX dephosphorylate an activated, mitogen-activated protein (MAP) kinase. The
XX activity, for modulation of a proliferative agents which modulate DSP-16
XX of a cell, or differentiation of a cell. The cell displays contact
XX inhibition of cell growth or anchorage independent growth and may display
XX altered intercellular adhesion. The agent may modulate apoptosis, or the
XX cell cycle. The identified modulators can be used to treat Duchenne
XX muscular dystrophy, cancer, graft-versus-host disease, autoimmune
XX diseases, allergies, metabolic diseases, abnormal cell growth, abnormal
XX cell proliferation, and cell cycle abnormalities. The present amino acid
XX sequence represents the human dual-specificity phosphatase-3 (DSP-16)
XX protein of the invention. This sequence is encoded by the human DSP-16
XX gene located on chromosome 12p
XX
XX Sequence 665 AA;
XX
XX Alignment Scores:
XX Pred. No.: 0 Length: 665
XX Score: 472.00 Matches: 672
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 70.76% Indels: 0
XX DB: 5 Gaps: 0
XX
XX US-10-029-345A-108_COPY_532_2532 (1-2001) x AAU79156 (1-665)
QY 586 ACCTGTCGAAGAGCTTATTCCTGAGTCTATTTCTGCTGCTGAGTGAATGAC 645
DB 194 ThrCysProIlySProAspPheIleProGlnSerHisPheLeuArgValProValAsnAsp 213
QY 646 ACCTTTTGTGAGAAATTTTGGCGGTGGAGCAAAATAGTAAATTTTATTTGAGAAAGCA 705
DB 214 SerPheCysGlnIlySLeuProIlyPheAspIlySerValAspPheIleGlnIlySAla 233
```

QY 706 AAAGCCTCAATGAGTGTCTAGTCACTGTTAGCTGGAGATCTCCCGTCGCACAC 765
 |||||
 Db 234 LyhlaSerhbnlgCyvValleuValHiScybleuHlaGlylleSerAhsSerAlaThr 253
 |||||
 QY 766 ATGCTATGCGCTTACATCATGAAGAGAGTGAACATGCTTTAGATGAAGCTTACAGATT 825
 |||||
 Db 254 HlaAlaIleAlaTyrIleMetLysArgMetLaspMetSerLeuHnbpGluAlaTyrArgPhe 273
 |||||
 QY 826 GTGAAGAAAAAGAACTACTATATCTCCAACTCAATTTTCTGGCCCAACTCTTGAGAC 885
 |||||
 Db 274 VallyleGluLysArgProThrIleSerProAsnPhenAsnPhenLeuGlyGlnLeuLeuAsp 293
 |||||
 QY 886 TATGAGAGAAGAATTAGAACCAGACTGAGACATCAAGGCCCAAGCAAACTCAAGCTG 945
 |||||
 Db 294 TyrGluLysLysIleLysAsnGlnThrGlyAlaSerGlyProLysSerLysLeuLysLeu 313
 |||||
 QY 946 CTGCACTGAGAGAACCAATGAACCTGTCCCTGTCTCAGAGGGGTGACAGAAAGC 1005
 |||||
 Db 314 LeuHileuGluLysProAsnGluProValProAlaValSerGluGlyGlnLysSer 333
 |||||
 QY 1006 GAGAGCGCCCTCAGTCCACCTGTGCGGACTCTGTCTACAGAGGACAGCAAAAG 1065
 |||||
 Db 334 GluThrProLeuSerProProCysAlaAspSerAlaThrSerGluAlaIleGlyGlnArg 353
 |||||
 QY 1066 CCGGTGATCCCGCAGCGTGCAGCGCTGCCAGCGTGCAGCGCTGTATGAGAGAC 1125
 |||||
 Db 354 ProValHisProAlaSerValProSerValProSerValGlnProSerLeuLeuGluAsp 373
 |||||
 QY 1126 AGCCCGCTGTATACAGCGCTCAGTGGGCTCAGCTGTCCGACAGACAGCTGGAAGACAG 1185
 |||||
 Db 374 SerProLeuValGlnAlaLeuSerGlyLeuHileuSerHlaAspArgLeuGluLysSer 393
 |||||
 QY 1186 AATTAAGCTCAAGGTTCTCTCTCTGTGATATCAATCAAGTTTCAATTCAGCCAGATG 1245
 |||||
 Db 394 AsnLysLeuLysArgSerPheSerLeuAspIleLysSerValSerTyrSerAlaSerMet 413
 |||||
 QY 1246 GCAGCATCTTACATGCTCTCTCTCATCAGAGAGATCTTGGAAATCTCAAAACCTTCC 1305
 |||||
 Db 414 AlaAlaSerLeuHileGlyPheSerSerSerGlnAspAlaLeuGluTyrTyrLysProSer 433
 |||||
 QY 1306 ACTACTCTGATGGAGCAACAAGCTATGCGAGTTCTCCCTGTTCAGAACTATCGAG 1365
 |||||
 Db 434 ThrThrLeuAspGlyThrAsnLysLeuCyGlnPheSerProValGlnLysLeuSerGlu 453
 |||||
 QY 1366 CAGACTCCGCAACCACTCTGATAGAGAGAGAGCCGACTCCCAAGAAAGCTGCAGACC 1425
 |||||
 Db 454 GlnThrProGluThrSerProAspLysGluGluAlaSerIleProLysLysLeuGlnThr 473
 |||||
 QY 1426 GCCAGGCTTACAGACGACGAGAGAGATGATCGGTCAGAAACACAGACAGTGGC 1485
 |||||
 Db 474 AlaArgProSerAspSerGlnSerLysArgLeuHileSerValArgThrSerSerSerGly 493
 |||||
 QY 1486 ACCGCCAGAGGTCCCTTTATCTCACTCATCGAAGTGGAGCGTGGAGAGCAATTAC 1545
 |||||
 Db 494 ThrAlaGlnArgSerLeuLeuSerProLeuHileArgSerGlySerValGluAspAsnTyr 513
 |||||
 QY 1546 CACACCAAGCTTCTTTTGGGCTTTCCACAGCCAGACAGACCTCAGCAAGTCTGTGGC 1605
 |||||
 Db 514 HisThrSerPheLeuPheGlyLeuSerThrSerGlnGlnHileuThrLysSerAlaGly 533
 |||||
 QY 1606 CTGGGCGCTTAAGGCTGGCACTGGGATATCTGGGCGCCGACAGCTTACCCCTCCCTG 1665
 |||||
 Db 534 LeuGlyLeuLysGlyTyrPheLysAspIleLeuAlaProGlnThrSerThrProSerLeu 553
 |||||
 QY 1666 ACCAGCAGCTGTATTTTGCACAGAGTCTTCACTTACTTACTGTGCTCAGCCATCTAC 1725
 |||||
 Db 554 ThrSerSerTyrTyrPheAlaThrGlnSerSerHisPheTyrSerAlaSerAlaIleTyr 573
 |||||
 QY 1726 GAGAGCGTGCAGTTACTCTGCTTACAGCTGCAGCCAGCTCCCACTTGGAGAGACAA 1785
 |||||
 Db 574 GlyGlySerAlaSerTyrSerAlaTyrSerCysSerGlnLeuProThrCysGlyAspGln 593
 |||||
 QY 1786 GTCTATTCTGTGGCGAGCGGCAAGACCAAGTACAGACTACTGCGCGGAGACTGG 1845
 |||||

Db 594 ValTyrSerValArgArgArgGlnLysProSerAspArgAlaAspSerArgAArgSerTyr 613
 |||||
 QY 1846 CATGAAGAGAGCCCTTTGAAAAAGCTTTAAACGCAAGAGCTGCCAAATGAAATTTGCA 1905
 |||||
 Db 614 HisGluGlnSerProPheGlnLysGlnPheLysArgArgSerCysGlnMetGluPheGly 633
 |||||
 QY 1906 GAGAGCATATGTCAGAGAAACAGTTCACGGGAAAGACTGGCGGAAAGGCGCACTGCTT 1965
 |||||
 Db 634 GluSerIleMetSerLysAsnArgSerArgGluGlnLeuGlyValGlySerGlnSer 653
 |||||
 QY 1966 AGCTTTTCGGGCGAGCATGGAATCATTTAGGCTTCC 2001
 |||||
 Db 654 SerPheSerGlySerMetGluIleIleGluValSer 665
 |||||
 RESULT 7
 ID AAU09946 standard; protein; 665 AA.
 XX
 AC AAU09946;
 XX
 DT 18-UTN-2002 (first entry)
 XX
 DE Protein sequence of human (dual specificity phosphatase) DUSP-10.
 XX
 KW Human; dual specificity phosphatase; DUSP-10; cancer; epilepsy; stroke;
 KW neuronal degeneration syndrome; Alzheimer's disease; depression;
 KW schizophrenia; asthma; immune disorder; inflammatory process; arthritis;
 KW osteoporosis; diabetes.
 XX
 OS Homo sapiens.
 XX
 PN WO200177340-A1.
 PD 18-OCT-2001.
 XX
 PF 06-APR-2001; 2001WO-BP003966.
 PR 10-APR-2000; 2000EP-00107143.
 XX
 PA (MERE) MERCK PATENT GMBH.
 XX
 PI Duecker K;
 XX
 DR WPI; 2002-010917/01.
 DR N-PSDB; AAS15768.
 XX
 PT Novel dual specificity phosphatase polypeptides useful for treating
 PT cancers, epilepsy, stroke, neuronal degeneration syndromes, Alzheimer's
 PT disease, depression, schizophrenia, asthma and immune disorders.
 XX
 PS Claim 2; Page 37-39; 43pp; English.
 XX
 CC The present invention relates to a new isolated dual specificity
 CC phosphatase (DUSP10) polypeptide, comprising a 665 residue amino acid
 CC sequence that is fully defined in the specification. The invention also
 CC provides a sequence encoded by a 3059 nucleotide sequence fully defined
 CC in the specification, and a sequence having at least 95 % identity to the
 CC polypeptide, or fragments or variants of DUSP-10. The invention is useful
 CC for treating cancer e.g. leukemia, colon carcinoma, lung cancer,
 CC prostate cancer, metastasis of tumour cells, neo-angiogenesis, epilepsy,
 CC stroke, neuronal degeneration syndromes, Alzheimer's disease, depression,
 CC schizophrenia, cardiac myopathies, asthma, immune disorders,
 CC inflammatory processes e.g. arthritis, bowel disease, type I diabetes,
 CC osteoporosis, diabetes and diabetes associated diseases. The molecules of
 CC the invention are also useful as vaccines for inducing immunological
 CC response in a mammal, in disease diagnosis and in assays for screening
 CC agonistic or antagonistic compounds. Other uses of the invention include
 CC identifying membrane bound or soluble receptors, as a diagnostic reagent,
 CC in chromosome localization studies, and as a valuable tool in tissue
 CC expression studies. The present sequence represents the dual specificity
 CC phosphatase, DUSP-10, protein of the invention
 XX

Db 634 GluSerIleWetSerGluSnaRgsSerArgGluGluLeuGlyIyValGlySerGlnSer 653
QY 1966 AGCTTTGGGGCAGCATGGAATCATTTGAGGTCTCC 2001
Db 654 SerPheSerClySerMetGluIleIleGluValSer 665
RESULT 9
ID ABB97946 standard; protein; 665 AA.
XX ABB97946;
AC ABB97946;
XX 06-SBP-2002 (first entry)
XX Human protein sequence #13.
DE Human protein sequence #13.
XX Human; brain; tonsil; hippocampus; foetal brain; diagnosis.
XX Homo sapiens.
OS Homo sapiens.
XX MO200252005-A1.
PN 04-JUL-2002.
XX 20-DEC-2001; 2001WO-JP011217.
PF 22-DEC-2000; 2000JP-00389742.
PR (KAZU-) KAZUSA DNA RES INST FOUND.
PA (CELE-) CELESTAR LEXICO-SCI LTD.
XX Ohara O, Nagase T, Nakajima D;
XX WPI: 2002-500762/53.
DR N-PSDB; ABB93966.
PT Genes and their expression products cloned from human cDNA libraries for
PT treatment and diagnosis of diseases associated with their expression.
XX
PS Claim 1(a); Page 112-116; 238pp; Japanese.
XX The invention relates to DNA encoding polypeptides directly cloned from
CC cDNA libraries originating in adult whole brain, human tonsil, human
CC adult hippocampus and human foetal brain. Polypeptides and
CC polynucleotides of the invention may be used in the investigation of
CC differential expression of the DNA sequences in normal subjects and
CC disease patients. They may also be used in the production of antibodies,
CC oligonucleotide probes and DNA chips for diagnosis and identification of
CC drugs for treatment of diseases with which the DNA sequences are
CC associated. The sequences given in records ABB97934-ABB97964 represent
CC human proteins of the invention
XX
SQ Sequence 665 AA;
Alignment Scores:
Pred. No.: 0 Length: 665
Score: 472.00 Matches: 472
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 70.76% Indels: 0
DB: 5 Gaps: 0
US-10-029-345A-108_COPY_532_2532 (1-2001) x ABB97946 (1-665)
QY 586 ACCGTGTCAGAGCTGACTTTATCCCGAGTCTCATTTCTCGTGCTGTGAATGAC 645
Db 194 ThrCysProLysProAspPheIleProGluSerHisPheLeuAlaGValProValAsnAsp 213
QY 646 AGCTTTTGAGAAAAATTTTGGCGTGTGGAACAATCAGTAGATTCTTGAGAAAGA 705
Db 214 SerPheCysGluLysIleLeuProThrLeuAspLysSerValAspPheIleGluLysAla 233
QY 706 AAGGCTCCAAATGATGTCTTCTAGTGACATGTTAGTGAGGATCTCCGCTCCGAC 765

Db 234 LysAlaSerAsnGlyCysValLeuValHisCysLeuAlaGlyIleSerArgSerIaThr 253
QY 766 ATCCGCTATCGCTATCATCATGAGAGATGACATGCTTTAGATGAAGCTTACAGATT 825
Db 254 IleAlaIleIaIaTyIleWetLysArgMetAspMetSerLeuAspGluAlaIaTyArgPhe 273
QY 826 GTGAAAGAAAAGAACCTTACTATCTCAAACTTCAATTTTCTGAGGCAATCTCGAGC 885
Db 274 ValLysGluLysArgProThrIleSerProAsnPheAsnPheLeuGlyGluLeuLeuAsp 293
QY 886 TATGAGAAAGAAATTAAAGAACCCAGACTGAGCATCAGGCGCCAAAGACCAATCAAGCTG 945
Db 294 TyGluLysLysIleLysAsnGlnThrGlyAlaSerGlyProLysSerLysLeuLysLeu 313
QY 946 CTGCACTCGGAGAACCCAAATGAACTCTGCTCCCTGCTGTCTCAGAGGCTGAGCAAAAAGC 1005
Db 314 LeuHisLeuGluLysProAsnGluProValProAlaValSerGluGlyGlyGluLysSer 333
QY 1006 GAGAGCGCCCTGAGTCCACCCCTGCGCACTCTGCTACCTCAGAGGAGGAGCAAAAG 1065
Db 334 GluThrProLeuSerProProCysAlaAspSerAlaThrSerGluAlaIaGlyIaThr 353
QY 1066 CCCGTGCAATCCCGCCAGCGTGCCAGCGTGCACCGTGCAGCCGCTGCTTTAGAGAC 1125
Db 354 ProValHisProAlaSerValProSerValProSerValGlnProSerLeuLeuGluAsp 373
QY 1126 AGCCCGCTGTACAGGCGCTCAGTGGCTGACCTGTCCGACAGACAGCTGGAACAACG 1185
Db 374 SerProLeuValGlnAlaLeuSerGlyLeuHisLeuSerAlaAspArgLeuGluAspSer 393
QY 1186 AATAAGCTCAAGCGTCTCTCTGTGATATCAATCAGTTTCAATATATACGCCAGATG 1245
Db 394 AsnLysLeuLysArgSerPheSerLeuAspIleLysSerValSerTyrSerAlaSerIle 413
QY 1246 GCAGCATCTTACATGAGCTTCTCTCATCAGAAAGATGCTTTGAACTACTACAACTTCC 1305
Db 414 AlaAlaSerLeuHisGlyPheSerSerSerGluAspAlaLeuGluTyrTyrLysProSer 433
QY 1306 ACTACTCTGAGTGGAGCAACAAGCTATGCCAGTTCTCCCTGTTCAGAACTATCGAG 1365
Db 434 ThrThrLeuAspGlyThrAsnLysLeuCysGlnPheSerProValGlnGluLeuSerIle 453
QY 1366 CAGACTCCCGAAACAGAGCTGTATGAGAGGAGGAGCAGATCCCAAGAGCTGACAGC 1425
Db 454 GlnThrProGluThrSerProAspLysGluGluAlaSerIleProLysLysLeuGlnThr 473
QY 1426 GCCAGGCTTGAAGACGACGACGAGCAAGGATTTGCAATTGGTACAAACGACGAGTGC 1485
Db 474 AlaArgProSerAspSerGlnSerLysArgLeuHisSerValArgThrSerSerGly 493
QY 1486 ACCGCCAGAGGTCCCTTTATCTGCATCTGCATGAAAGTGGAGCGGTGAGACAAATTAC 1545
Db 494 ThrAlaGlnArgSerLeuLeuSerProLeuHisArgSerGlySerValGluAspAsnLys 513
QY 1546 CACACAGACTCTTTTGGGCTTCCACAGACGACGACGACCTCACAGAGCTGTGAGC 1605
Db 514 HisThrSerPheLeuPheGlyLeuSerThrSerGlnGlnHisLeuThrLysSerAlaGly 533
QY 1606 CTGGGCTTTAAGGGCTGACACTCGATATCTTTGGGCCCCCAGACCTTACCTTCCCTG 1665
Db 534 LeuGlyLeuLysGlyTyrPheHisSerAspIleLeuAlaProGlnThrThrProSerLeu 553
QY 1666 ACCAGCAGCTGTATTTTGGCAGACAGTCTTCACTTCACTTCTGCTCAGCCATCTAC 1725
Db 554 ThrSerSerTyrPyrPheAlaThrGluSerSerHisPheTyrSerAlaSerAlaIleTyr 573
QY 1726 GAGGCAAGTGCAGTTACTGCTCAGCTGACGACGACGACGACGACGACGACGACGACGAC 1785
Db 574 GlyLysSerAlaSerTyrSerAlaTyrSerCysSerGlnLysProThrCysGlyAspGln 593
QY 1786 GTCTATTCTGTGCGCAGGCGGACAGAGCAAGTACAGAGCTGACTCGCGGAGCTGG 1845

Db 594 ValTyrSerValArgArgGlnLysProSerAspArgAlaAspSerArgArgSerTyr 613
QY 1846 CATGAAGAGAGCCCTTTGAAAAGCACTTAAACGACGAGCTGCCAAATGGAATTGCA 1905
Db 614 HAsGluGlnSerProPheGlnLysGlnPheLysArgArgSerCysGlnMetGlnPheGly 633
QY 1906 GAGAGCATCTGTCAGAGAACAGAGTCAACGGAAGAGCTGGGAAAGTGGCACTGCT 1965
Db 634 GluSerIleMetSerGluMetArgSerArgGlnGlnLeuGlnValGlySerGlnSer 653
QY 1966 AGCTTTTCGGGACGACATGGAATCATTTAGAGTCTCC 2001
Db 654 SerPheSerGlySerMetGluIleIleGlnValSer 665
RESULT 10
AAU79929
ID AAU79929 standard; protein; 665 AA.
AC AAU79929;
XX
DT 02-JUL-2002. (first entry)
XX
DE Human dual specificity phosphatase 21117 protein.
XX
KM Human: dual specificity phosphatase 21117; erythroid-related disorder;
KM haematopoietic-related disorder; leukaemia; autoimmune disease; anaemia;
KM erythrocytosis; liver-related disorder; cancer.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Domain 11..131
FT 11..131 /label= Rhodanese_like_domain
FT 21..24 /note= "Casein kinase II phosphorylation site"
FT 91..94 /note= "Casein kinase II phosphorylation site"
FT 158..297 /note= "Casein kinase II phosphorylation site"
FT 214..217 /note= "Dual specificity phosphatase catalytic domain"
FT 242..254 /note= "Casein kinase II phosphorylation site"
FT 242..254 /note= "Tyrosine specific protein phosphatase active site"
FT 242..254
FT 265..269 /note= "C-X5-R motif"
FT 369..372 /note= "Casein kinase II phosphorylation site"
FT 421..424 /note= "Casein kinase II phosphorylation site"
FT 434..437 /note= "Casein kinase II phosphorylation site"
FT 458..461 /note= "Casein kinase II phosphorylation site"
FT 508..511 /note= "Casein kinase II phosphorylation site"
FT 589..592 /note= "Casein kinase II phosphorylation site"
FT 612..615 /note= "Casein kinase II phosphorylation site"
FT 617..620 /note= "Casein kinase II phosphorylation site"
FT 642..645 /note= "Casein kinase II phosphorylation site"
FT 642..645 /note= "Casein kinase II phosphorylation site"
XX
PN US2002034807-A1.
XX
PD 21-MAR-2002.
XX
PF 23-MAR-2001; 2001US-00816494.

PR 24-MAR-2000; 2000US-0191858P.
XX
PA (MEYE/) MEYERS R A.
XX
PI Meyers RA;
XX
DR WPI; 2002-351088/38.
XX
DR N-PSDB; ABK49402.
XX
PT New nucleic acids, designated 38692 and 21117, encoding dual specificity
PT phosphatases for treating cell proliferation and differentiation
PT disorders including hematopoietic and erythroid-related disorders and
PT cancers.
XX
PS Claim 8; Fig 1; 76bp; English.
XX
CC The present invention relates to new nucleic acids designated 38692 and
CC 21117 encoding dual specificity phosphatase family members. The nucleic
CC acid, polypeptide encoded by it, and antibody specific for the
CC polypeptide may be used to diagnose and treat hematopoietic-related
CC disorders such as leukemias and autoimmune diseases, erythroid-related
CC disorders such as anemias and erythrocytosis, liver-related disorders,
CC and cancers, particularly of the breast, colon, adipose, prostate and
CC lung. The present amino acid sequence represents the human dual
CC specificity phosphatase 21117 protein of the invention, as described
CC above
XX
SQ Sequence 665 AA;
XX
Alignment Scores:
Pred. No.: 0 Length: 665
Score: 472.00 Matches: 472
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 70.76% Indels: 0
DB: 5 Gaps: 0
US-10-029-345A-108_COPY_532_2532 (1-2001) x AAU79929 (1-665)
QY 586 ACCTGTCGAAGAGCTGATTTATCCCGAGTCTGATTTCTCGTGGCTGTGAATGAC 645
Db 194 ThrCysProLysPheProPheIleProGlnSerHisPheLeuArgValProValAsnAsp 213
QY 646 AGCTTTTGAGAAATTTTGGCGGTGGACAAATCAGTAGATTTCATTGGAAGCA 705
Db 214 SerPheCysGlnLysIleLeuProTyrLeuAspLysSerValAspPheIleGlnLysVal 223
QY 706 AAAGCTCCAAAGAGTGTCTGAGTCACTGTTAGCTGGGATCCCGCTCGGCACC 765
Db 234 LysAlaSerAsnGlnCysValLeuValHisCysLeuAlaGlyIleSerArgSerAlaThr 253
QY 766 ATCGATATGCGCTTACATCATGAAGAGATGACATGCTTTAGATGAAGCTTACAGATT 835
Db 254 IleAlaIleAlaTyrIleMetLysArgMetAspMetSerLeuAspGlnAlaTyrArgPhe 273
QY 826 GTGAAGAAGAAAGACCTATATCTCCAACTTCAATTTTGGGCCAACTCTGGAC 885
Db 274 ValLysGlnLysArgProThrIleSerProAsnPheAsnPheLeuGlnLeuLeuAsp 293
QY 886 TTGGAAGAAGAGTTTGAAGAACCGAGTGAAGATCAAGGGCCAAAGCAAACTCAAGCTG 945
Db 294 TyrGlnLysValLysValAsnGlnThrGlyAlaSerGlyProLysSerLysLeuLysLeu 313
QY 946 CTGACCTGAGAGAGCAATGAATGAACCTGCTGCTGCTGCTGACAGAGGTGACAGAAAGC 1005
Db 314 LeuHisLeuGlnLysArgAsnGlnLysProValProAlaValSerGlnLysGlnLysSer 333
QY 1006 GAGACGCGCTCAATGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1065
Db 334 GluThrProLeuSerProProCysAlaAspSerAlaThrSerGlnLysAlaGlyGlnArg 353
QY 1066 CCGGTGCATCCGCGCGCGCTGCGCGAGCGTCCAGCGTCCAGCGCTGCTGAGAGGAC 1125

Db 354 ProValHisProAlaSerValProSerValProSerValGlnProSerLeuLeuGlnAsp 373
QY 1126 AGCCCGGTGTACAGGCGCTCACTGGGCTGCACCTGTCCGACAGACAGCTGGAAACAGC 1185
Db 374 SerProLeuValGlnAlaLeuSerGlyLeuHisLeuSerAlaAspArgLeuGlnAspSer 393
QY 1186 AATAAGCTCAAGGGTCCCTCTCTGTGATATCAATTCAGTTTCATATTCAGCCACGATG 1245
Db 394 AsnLysLeuLysArgSerPheSerLeuAspLysLeuLysSerValSerLysSerAlaSerMet 413
QY 1246 GCAGCATCCTTACATGGCTTCTCCTCATCAGAAATGCTTTGGAAATACATAAACTTCC 1305
Db 414 AlaAlaSerLeuHisGlyPheSerSerSerGluAspAlaLeuGlnLysTrpLysProSer 433
QY 1306 ACTACTCTGATGGACCAACAAAGCTATGCGAGTTCTCCCTGTTCAAGAACTATGGAG 1365
Db 434 ThrThrLeuAspLysLysThrAsnLysLeuGlyGlnPheSerProValGlnLeuSerGln 453
QY 1366 CAGACTCCCGAAACCGATCCTGATTAAGAGGAGGAGCCAGCATCCCGCAAGAACTGGAGCC 1425
Db 454 GlnThrProGlnThrSerProAspLysGlnGlnAlaSerLysProLysLysLeuGlnThr 473
QY 1426 GCCAGGCTTACAGACAGCCAGACCAAGCATTCATTCGGTCAAGAACAGACAGCTGGC 1485
Db 474 AlaArgProSerAspSerGlnSerLysArgLeuHisSerValArgThrSerSerSerGly 493
QY 1486 ACCGCCCAAGAGTCCCTTTATCTCCATCTGCAATCGAAGTGGAGCGTGGAGCAATTAC 1545
Db 494 ThrAlaGlnArgSerLeuSerProLeuHisArgSerLysSerValGlnAspAsnLys 513
QY 1546 CACACAGACTCTCTTTTCGGCTTTTCCACAGCCAGCAGACAGCACTCCACGAATGCTGCTGC 1605
Db 514 HisThrSerPheLeuHeuHeGlyLeuSerThrSerGlnGlnHisLeuThrLysSerAlaGly 533
QY 1606 CTGGGCGCTTAAGGGCTGGCACTGGAATCTTGGCCCCCAGACCTTACCCCTTCCCTG 1665
Db 534 LeuGlyLeuLysGlyLysThrPheHisSerAspLysLeuAlaProGlnThrSerThrProSerLeu 553
QY 1666 ACCAGAGCTGGATTTTGGCAAGAGTCTCTACACTTCTACTCTGCTCAGCCATCTAC 1725
Db 554 ThrSerSerTrpLysPheAlaThrGlnSerSerHisPheLysSerAlaSerAlaLysTrp 573
QY 1726 GGAGCAGTGCAGCTTACTGCTGCTACAGCTGACGAGCTGGCCACTGGCGAGACCA 1785
Db 574 GlyGlySerAlaSerLysTrpSerAlaLysTrpSerCysSerGlnLeuProThrCysGlyAspGln 593
QY 1786 GTCTATTCTGTGGCAGCGCGGAGCAAGCAAGTACAGAGCTGACTCGCGCGGAGCTGG 1845
Db 594 ValLysSerValArgArgArgGlnLysProSerAspArgAlaAspSerArgArgSerTrp 613
QY 1846 CATGAAGAGAGCCCTTTGAAAAGCAGTTTAAACGCAAGAAAGCTGCCAATGGAATTGGA 1905
Db 614 HisGlnGlnSerProHeuHeGlyLysGlnPheLysArgArgSerCysGlnMetGlnPheGly 633
QY 1906 GAGAGCATATGTCAGAGAAACAGGTACAGGAGAGACTGGGAGAAAGTGGGAGTCAAGTCT 1965
Db 634 GluSerLysLeuMetSerLysAsnArgSerArgGlnGlnLeuGlyLysValGlyLysGlnSer 653
QY 1966 AGCTTTTGGGAGCATGGAATCATTTGAGTCTCC 2001
Db 654 SerPheSerGlySerMetGlnLysLeuValSer 665

RESULT 11
ID ABB97291
ID ABB97291 standard; protein; 665 AA.
XX
AC ABB97291;
XX
DT 27-JUN-2002 (first entry)
XX
DE Novel human protein SEQ ID NO: 559.
XX
KW Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;

KW antinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW neuroprotective; antiparkinsonian; protein therapy; EST;
KW expressed sequence tag.
XX
OS Homo sapiens.
XX
PN MO200222660-A2.
XX
PD 21-MAR-2002.
XX
PF 10-SEP-2001; 2001WO-US026015.
XX
PR 11-SEP-2000; 2000US-00659671.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
DR WPI; 2002-292408/33.
XX
DR N-PSDB; ABN32477.
XX
PT An isolated polynucleotide for treating diseases associated with its
PT encoded polypeptide such as cancer and multiple sclerosis.
XX
PS Example 2; SEQ ID NO 559; 509bp; English.
XX
CC The present invention provides the protein and coding sequences of 444
CC novel human proteins. These were isolated from expressed sequences tags
CC (ESTs). They can be used to stimulate cell growth, to regulate
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC e.g. in burn treatment, to regulate the immune system e.g. to treat
CC multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke
CC and cancer, to screen for drugs, to treat inflammatory conditions e.g.
CC rheumatoid arthritis, and to treat nervous system disorders e.g.
CC Parkinson's disease. The present sequence is a protein of the invention
XX
SQ Sequence 665 AA:
XX
Alignment Scores:
Pred. No.: 0 Length: 665
Score: 472.00 Matches: 472
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 70.76% Indels: 0
DB: 5 Gaps: 0
US-10-029-345A-108_COPY_532_2532 (1-2001) x ABB97291 (1-665)
QY 586 ACCTGTCCAAAGCCGACTTATCCCGAGTCTATTCCTGCGCGTGCCTGAATGAC 645
Db 194 ThrCysProLysProAspPheLeuProGlnSerHisPheLeuArgValProValAsnAsp 213
QY 646 ACCTTTTGGAGAAATTTTGGCGTGTGGCAATAGTATGATTTTATTGAGAAACCA 705
Db 214 SerPheCysGlnLysLysLeuProThrLeuAspLysSerValAspPheLysGlnLysAla 233
QY 706 AAAGCTCCCAATGATGTGTTCTAGTGACATGTTAGCTGGGATCTCCGCTCGCCACC 765
Db 234 LysAlaSerAsnGlyCysValLeuValHisCysLeuAlaGlyLysSerArgSerAlaThr 253
QY 766 ATCGCTATCGCTTAATCATGAAGAGTATGACATGCTTTAGATGAAGCTTACGATTT 825
Db 254 IleAlaIleAlaLysLysLeuMetLysArgMetAspMetSerLeuAspGlnAlaLysArgPhe 273
QY 826 GTGAAGAAAAAAGACTACTATATCTCCAAACTCAATTTTCTGGGCAACTCTGGAC 885
Db 274 ValLysGlnLysArgProThrThrLysSerProAsnPheAsnPheLeuGlyGlnLeuLysAsp 293
QY 886 TATGAGAAAGATTAAGAACCAAGCTGAGCATCAGGCGCAAGAGCAAACTCAAGCTG 945
Db 294 TyrGlnLysLysLysLysLeuAsnGlnThrGlyLysSerGlyProLysSerLysLeuLysLeu 313

```

QY 946 CTGCACCTGGAGAGCAAAATGAACCTGCTCCCTGCTCTCAGAGGGGTGAGCAAGAAAAGC 1005
DB 314 LeuH1leucluluySPRoahngluProvalProh1valsergluglygluluySer 333
QY 1006 GAGACGGCCCTCAGTCCAGCCCTGTGCGACTGTGCTACCTCAGAGGCGAGGAGCAAAAGG 1065
DB 334 GluTherProleuserProProCyb1aAspSerlathSerglunla1aaglyglunArg 353
QY 1066 CCCGTGATCCCGCCAGCCGTGCCAGGTGCCAGCTGACGCCGTGACGCCGTGTTAGAGGAC 1125
DB 354 ProvalH1aProh1aSerValProSerValGlnProSerleuLeuGluAsp 373
QY 1126 AGCCCGCTGGATCAGCGGCTCAGTGGGCTGACCTGCCGAGACAGCTGAGAGACAGC 1185
DB 374 SerProleuValGln1aLeuSerGlyLeuH1leuSerl1aAspH1gLeuGluAspSer 393
QY 1186 AATAAGCTCAAGCGTTCCTTCTCTGTGATATCAAAATCAGTTTCATATTCAGCCAGCATG 1245
DB 394 AsnlySerleuylAspSerPheSerleuAsp11elySerValSerTySerl1aSerMet 413
QY 1246 GCAGCATCTTACATGGCTTCTCTCATCAGAAATGCTTTGGAATCTCAAACTTCC 1305
DB 414 Ala1aSerleuH1sglyPheSerSerSerGluAsp1aLeuGluTyTyTyPProSer 433
QY 1306 ACTACTGTGATGGAGCAACAAGCTATGCGAGTTCGCCCTGTTCAGAACTATGCGAG 1365
DB 434 ThrTherleuAspGlyH1nAsnlyLeuCybglInPheSerProvalGlnGluLeuSerGlu 453
QY 1366 CAGACTCCCGAAAACAAGTCTGTATAGAGAGAAAGCCAGCATCCCAAGAAAGTGCAGACC 1425
DB 454 GlnTherProglunTherSerProAspLygluglu1aSerl1eProlyblybLeuGlnThr 473
QY 1426 GCCAGGCTTTCAGACAGCCAGAGCAAGCGATTGCGTCAAGAACCCAGCAGCAGTGC 1485
DB 474 AlaArgProSerAspSerGlnSerlyArgLeuH1leuSerValArgTherSerSerGly 493
QY 1486 ACCGCCAGAGGTCCTTTATCTCCATCGATCGAAGTGGGCGCGGAGGAGCAATTCAC 1545
DB 494 Thr1aGln1aArgSerleuLeuSerProleuH1eArgSerl1ySerValGluAspH1Ty 513
QY 1546 CACACACAGCTTCCTTTCCGCTTTCCACAGCCAGGAGGAGCACTCAAGAACTGCTGC 1605
DB 514 HisTherSerheuleuPhegllyLeuSerThSerGlnlnH1leuThrybSerl1aGly 533
QY 1606 CTGGGCTTAAAGGCTGCGCACTGGATATCTTGCCCCCAAGCTTACCCCTTCCCTG 1665
DB 534 LeuGlyLeuylsglyTrpH1SerAsp11eLeuAlaProGlnTherSerThProSerLeu 553
QY 1666 ACCGACAGCTGATATTTTGCACAGAGTCTCCACACTTCACTGCTGACGATCTAC 1725
DB 554 ThSerSerTyPyrPheAlaThrGlnUserSerThSerl1eH1ySerl1aSerl1a1eTy 573
QY 1726 GAGAGCAGTCCAGATTACTGTGCTACAGCTGACGAGCAGCTGCGCACTTCGAGAGACAA 1785
DB 574 GlyGlySerl1aSerTySerl1aTySerCySerGlnleuProThrCybgllyAspGln 593
QY 1786 GTCTATTCTGCGCAGCGGCGAGAAAGCAAGTGAAGAGCTGACTGCGCGAGAGCTGG 1845
DB 594 ValTySerValArgArgArgGlnlyPProSerAspArg1aAspSerArgAgSerSerTrp 613
QY 1846 CATGAAGAGAGCCCTTTGAAAAGCACTTTAAAGCAAGAGCTGCCAAATGAAATTTGCA 1905
DB 614 HisGlnGlnUserProPheGlnlybGlnPheLyAspArgSerCybGlnMetGluPheGly 633
QY 1906 GAGAGCATCATGTCAAGAGACAGGTCAAGGAGAGAGCTGGGAGAAAGTGGCAGTCACT 1965
DB 634 GluSerl1eMetSerGluAsnArgSerArgGluGluLeuGlyValGlySerGlnSer 653
QY 1966 AGCTTTTGGGCGAGCATGAAATCATTTAGAGTCTCC 2001
DB 654 SerPheSerGlySerMetGlu1le1leGluValSer 665

```

```

RESULT 12
ID ADA54744
AC ADA54744;
DT 20-NOV-2003 (first entry)
DE Human protein, SEQ ID 2312.
KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
KW Inflammatory disease; osteoporosis; neurological disease.
OS Homo sapiens.
PN EP1293569-A2.
PD 19-MAR-2003.
PF 21-MAR-2002; 2002EP-0006586.
PR 14-SEP-2001; 2001JP-00328381.
PR 24-JAN-2002; 2002US-0350435P.
XX
XX (HELI-) RES ASSOC BIOTECHNOLOGY.
PA (REAS-) RES ASSOC BIOTECHNOLOGY.
PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,
PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,
PI Seki N, Yoshikawa T, Otsuka M, Nagahara K, Masuno Y;
DR WPI, 2003-395539/38.
DR N-PDB; ADA53105.
XX
XX New polynucleotides encoding full-length polypeptides, e.g. secretory in
PT and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.
XX
XX Claim 14; SEQ ID NO 2312; 205pp; English.
XX
XX The present invention relates to novel human secretory or membrane
XX proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
XX ADA54072). The coding sequences are useful in the gene therapy of
XX diseases caused by abnormalities of the proteins, e.g. cancer,
XX inflammatory diseases, osteoporosis or neurological disease.
XX
XX Sequence 665 AA;
SQ
Alignment Scores:
Pred. No.: 0 Length: 665
Score: 472.00 Matches: 472
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 70.76% Indels: 0
DB: 6 Gaps: 0
US-10-029-345A-108_COPY_532_2532 (1-2001) x ADA54744 (1-665)
QY 586 ACCGTGCAAAAGCTGACTTTATCCCGAGTGCATTTCTGCGTGGCTGGAATGAC 645
DB 194 ThrCybProlySerProAspPhe1leProGlnUserH1aPheLeuArgValProvalAsp 213
QY 646 AGCTTTTGGAGAAATTTTGCCTGTGGAACAAATCAGTATTTCAATTGAGAAAGCA 705
DB 214 SerPheCyGluLyserleuProTrpLeuAspLySerValAspH1eGluLybAla 233
QY 706 AAAGCTCCAAATGATGTGTCTTAGTGCACCTGTTAGCTGGGATCTCCGCTCCGCAAC 765
DB 234 LybAlaSerAsnGlyCybValLeuValH1aCybLeuAlaGlyl1eSerArgSerl1aThr 253
QY 766 ATCGCTATGCGCTACATCATGAAGAGGATGAGCATGCTTTTGATGAAGCTTACAGATT 825

```

Db	254	l l e a l l e a l t a r y l l e m e l y s a r g m e c a s p m e t e r l e u m s p g l u a l a r y a r g p h e	273
Qy	826	g t g a a g a a a a a a g a c c t a c t a t a t c t c c a a c t t c a a t t t t c t g g c c c a a c t c t g g a c	885
Db	274	v a l l y s g l i u y s a r p r o t h r l l e s e r p r o a s n p h e a n p h e u g l y g l n l e u l e u s p	293
Qy	886	t a t g a g a a a a g a t t a a g a c c a g a c t g g a g a c a t c a g g g c c a a a g a g a c c a a c t c a g c t g	945
Db	294	t y r g l u y s l y s l l e l y s a s n g l i n t h e l y a l a s e r g l y p r o l y s s e r l y b l e u l y b l e u	313
Qy	946	c t g c a c c t g g a g a a c c a a a t g a a c t g t c c c t g c t g t c a g a g g t g a c a g a a a a g c	1005
Db	314	l e u h i s l e u g l i u y s p r o a e n g l i u y p r o a l a l a l a s e r g l u g l y g l n l y s e r	333
Qy	1006	g a g a c g c c c c t c a g t c c a c c c t g g c c g a c t c t g c t a c c t c a g a g g a g a g a c a a a g	1065
Db	334	g l u t h r p r o l e u s e r p r o p r o c y a l a s p e r l a t h r s e r g l u a l a a g l y g l n a r g	353
Qy	1066	c c c g t g a c t c c c g c a g c g t c c c a g a g c t g c c a g c g t g a c g c c g t g t t a g a g a c	1125
Db	354	p r o l a l i n s p r o l a s e r v a l p r o s e r v a l p r o s e r v a l g l n p r o s e r l e u l e u g l u a s p	373
Qy	1126	a g c c g c t g t a c a g c g c t c a g t a g t g g c t g c a c t c t g c c g a c a c a g c t g g a a g a c a c	1185
Db	374	s e r p r o l e u v a l g l n a l a l e u s e r g l y l e u h i s l e u s e r l a a b a r g l e u g l u a s p s e r	393
Qy	1186	a a t a a g c t c a g a c g t t c c t c t c t g g a t a t c a a t c a g t t c a t a t t c a g c c a g a t g	1245
Db	394	a s n l y s b l e u y s a r g s e r p h e s e r l e u a s p r l e l y s s e r v a l s e r l y s e r l a s e r m e t	413
Qy	1246	g c a g c a t c c t t a c a t g a c t t c c t c a t c a g a a g a t c t t t g g a a t a c t a c a a c t t t c c	1305
Db	414	a l a l a s e r l e u h i l e g l y p h e s e r s e r g e r g l n s p r a l e u g l u y r t y l y s p r o s e r	433
Qy	1306	a c t a c t c t g a r t g g s a c c a a c a a g a t a g c c a g t t c c c t g t t c a g a a c t a t t c g a g	1365
Db	434	t h r t h r l e u a s p g l y t h r a s n l y s e u y s e l n p h e s e r p r o v a l g l n g l u l e u s e r g l u	453
Qy	1366	c a g a c t c c c g a a a c c a g t c c t g a t a a g a g a a g c c a g a c a t c c c a a a g c t g c a a c c	1425
Db	454	g l n t h r p r o g l u t h r s e r p r o a s p r y s g l u a l a s e r l l e p o l y s l y l e u g l n t h r	473
Qy	1426	g c c a g c c t t c a g a a g c c a g a c a a g a g a t t g a t t g g c r c a g a a c c a g a a g a g g c	1485
Db	474	a l a a r g p r o s e r a s p e r s e r l i n s e r l y a r g l e u h i s s e r v a l a r g l n s e r s e r g l y	493
Qy	1486	a c c g c c c a g a g g t c c c t t t a t c t c a c t g c a t c a a a t g g a g c g t g a g a c a a t t a c	1545
Db	494	t h r l a g l n a r g s e r l e u e u s e r r o l e u h i s a r g s e r g l y s e r v a l g l u a s p a n t y r	513
Qy	1546	c a c a c a a g c t t c t t t t t g g c c t t t c c a c a c a c c a g c a c c t a c a a a g t c t g c t g g c	1605
Db	514	h i s t r i s e r p h e l e u p h e g l y l e u s e r t h r s e r g l n g l n h l e n t h l y s e r l a e l y	533
Qy	1606	c t g g g c c t t a a g g c g g c a c t c g a g a t a t c t t g g c c c c c a g a c c t a c c c t c c c t g	1665
Db	534	l e u g l y l e u l y g e l y t t p h i s s e r s p l l e u a l a p r o g l n t h s e r t h r p r o s e r l e u	553
Qy	1666	a c c a g a c a g t g a t t a t t t t t g c a c a g a g t c c t a c a c t t a c t g c c t c a g c a c a t t a c	1725
Db	554	t h r s e r s e r t r p l y p h e a l a t h r g l u s e r s e r h i s p h e t y s e r l a s e r l a l l e t y r	573
Qy	1726	g g a g c a g r g c a g t t a c t c t g c c t a a c a g c t g a c a c a g c c c a c a t t g g c g a a c a a	1785
Db	574	g l y g l y s e r l a s e r t y s e r l a t y s e r c y s s e r g l n l e u p r o t h r c y e l y a s p o l n	593
Qy	1786	g t c t a t t c t g t c g c a g c a g c g a g a a c c a a g t g a c a g a c t a c t g c g g c g a g c t g g	1845
Db	594	v a l t y r s e r v a l a r g l a r g l n l y s p r o s e r a b a r g l a a s p e r a r g a r g e r t t r p	613
Qy	1846	c a t g a a g a a g c c c t t t a a a a g c a g t t a a c c a a a a g c t g c c a a t g a a t g a a t t t g a	1905
Db	614	h s a g l u g l u s e r p r o p h e g l i u y s g l n p h e l y s a r g a r g s e r c y g l m e c l i n h e g l y	633

QY	1906	GAGAGCATCATGTTCAGAGAACAGGTCA	CGGGAGAGGCTGGGGAAGTGGCAGTCACTCT	1965
DB	634	GIberilIeMetsSerGIuAenAArgSerA	rgIuIuenuGIySValGIySerGIInsEr	653
QY	1966	AGCTTTTCGGGACAGCATGGAATCAT	TATTGAGGTCTCC	2001
DB	654	SerPheserGIySerMeGIuIleIleGI	uValSer	665
RESULT 13				
ID	AAU79161	standard; protein: 665 AA.		
XX	AAU79161;			
AC	02-JUL-2002	(first entry)		
XX				
DE	Human dual-specificity phosphatase-3 (DSP-16) mutant protein #1.			
XX				
KW	Human; dual-specificity phosphatase-3; DSP-16; autoimmune disease;			
KW	mitogen-activated protein kinase; MAP; Duchenne muscular dystrophy;			
KW	cancer; graft-versus-host disease; allergy; metabolic disease;			
KW	abnormal cell growth; abnormal cell proliferation; contact inhibition;			
KW	cell cycle abnormality; anchorage independent cell growth; apoptosis;			
KW	intercellular adhesion; DSP-16 modulator; mutant; munein.			
XX				
OS	Homo sapiens.			
OS	Synthetic.			
FH	Key	Location/Qualifiers		
FT	Misc-difference 213	/note= "Wild-type Asp substituted by Ala"		
XX				
XX	WO200226997-A2.			
XX	04-APR-2002.			
XX	25-SEP-2001; 2001MO-US030124.			
PR	26-SEP-2000; 2000US-0235487P.			
PA	(CEPT-) CEPTYR INC.			
PI	Luche RM, Wei B;			
XX				
XX	WPI; 2002-315802/35.			
XX				
PT	New DSP-16 polypeptide, useful for identifying modulators of its			
PT	activity, which can be used in the treatment of disorders such as			
PT	Duchenne muscular dystrophy, or cancer.			
PS	Claim 46; Page: 87pp; English.			
XX				
XX	The present invention relates to a new polypeptide, DSP-16, having a 665			
CC	amino acid sequence, given in the specification, or a variant having at			
CC	least 50 % identical residues, which retains the ability to			
CC	dephosphorylate an activated mitogen-activated protein (MAP) kinase. The			
CC	invention can be used for identifying agents which modulate DSP-16			
CC	activity, for modulation of a proliferative response in a cell, survival			
CC	of a cell, or differentiation of a cell. The cell displays contact			
CC	inhibition of cell growth or anchorage independent growth and may display			
CC	altered intercellular adhesion. The agent may modulate apoptosis, or the			
CC	cell cycle. The identified modulators can be used to treat Duchenne			
CC	muscular dystrophy, cancer, graft-versus-host disease, autoimmune			
CC	diseases, allergies, metabolic diseases, abnormal cell growth, abnormal			
CC	cell proliferation, and cell cycle abnormalities. The present amino acid			
CC	sequence represents the human dual-specificity phosphatase-3 (DSP-16)			
CC	mutant protein #1. Note: This sequence is not shown in the specification			
CC	but is derived from the wild-type human DSP-16 (AAU79156) protein given			
XX	in figure 2 of the specification			
XX				
XX	Sequence 665 AA;			

Alignment Scores:

Pred. No.: 0
 Score: 452.00
 Percent Similarity: 100.00%
 Best Local Similarity: 100.00%
 Query Match: 67.77%
 DB: 5
 Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x AAU79161 (1-665)

```

QY 646 AGCTTTTGAGAAATTTTCCGCTGGTGAACAATCAGTAGATTTCATTGAGAAAGCA 705
DB 214 SerPheCyseGluysrIleuProtrPleuAspIysSerValAspPheIleGluysrAla 233
QY 706 AAGGCTCCCAATGAGTGTCTTCTAGTCCACTGTTTACCTGGAGTCCCGCTCCGACACC 765
DB 234 LysAlSerAsnGlyCyValLeuValHisCyLeuValAGlyIleSerAlaGSerAlaThr 253
QY 766 ATGCTATCCGCTACATCATGAGAGAGATGACATGCTTTAGATGAGCTTACAGATT 825
DB 254 ILeAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAlaIleAla 273
QY 826 GTGAAAGAAAAAGACTTACTATATCTCAAACTTCAATTTCTGGCCCACTCTGAC 885
DB 274 ValIyGluysrArgProthrIleSerProAsnPheAsnPheLeuGluIleuLeuAsp 293
QY 886 TATGAGAAAGAAATTAAAGAACAGACTGAGCATCAGAGGCCAAAGCAAACTCAAGCTG 945
DB 294 TyrGluysrIleysrIleysrAsnGluIleGlyAlaSerGlyProIysSerIleuysrLeu 313
QY 946 CTGACCTGGAGAGAACCAATGAACCTGCTCTGCTGTCTCAGAGGGTGGACAGAAAGC 1005
DB 314 LeuHisLeuGluysrProAsnGluProvalProAlaValSerGluGlyGluysr 333
QY 1006 GAGACGCCCTCAGTCACCTGTGCGACTGTCTACTCAGAGGACGAGGACAAAG 1065
DB 334 GluThrProLeuSerProProCyAlaAspSerAlaThrSerGluAlaIleGlyGlnArg 353
QY 1066 CCCGTGATCCCCCGACCGTCCCAAGGTGCCAGGCTGAGCGCTGCTTAAAGAGAC 1125
DB 354 ProValHisProAlaSerValProSerValGlnProSerLeuLeuGluAsp 373
QY 1126 AGCCGCTGTACAGGCGCTCAGTGGGCTGACCTGCGCAGACAGGCTGAGAGACAG 1185
DB 374 SerProLeuValGlnAlaLeuSerGlyLeuHisLeuSerAlaAspArgLeuGluAspSer 393
QY 1186 AATAAGCTCAAGGCTTCTCTCTGTGATATCAAACTCAGTTTCATTATTCAGCCAGCAT 1245
DB 394 AsnIleuysrArgSerPheSerLeuAspIleIysSerValSerIleSerAlaSerMet 413
QY 1246 GCGACATCTTACATGGCTTCTCTCATCAAGAGATGCTTGAATPCTAACAACTTCC 1305
DB 414 AlaAlaSerLeuHisGlyPheSerSerSerGluAspAlaLeuGluIleGlyIleProSer 433
QY 1306 ACTACTGTGATGGAGCAACAAGATATGACAGTTCCTCCCTGTACAGAACTATGGAG 1365
DB 434 ThrThrLeuAspIleGlyHisAsnIleGlyCysGlnPheSerProValGlnIleuSerGln 453
QY 1366 CAGACTCCCGAAAACAAGTCTCTGATTAAGAGAGAACAGATCCCAAGAGCTGACAGAC 1425
DB 454 GlnThrProGluThrSerProAspIleGluGluAlaSerIleProIleGlyIleuGlnThr 473
QY 1426 GCGAGGCTTCAAGACGCGCAGAGACGATTTGCTGCTGCAAGACGACAGAGTGGC 1485
DB 474 AlaArgProSerAspSerGlnSerIleArgLeuHisSerValArgThrSerSerSerGly 493
QY 1486 ACCGCCAGAGAGTCCCTTTATCTCCATCATCAAGTGGAGCGTGGAGGACCAATTAC 1545
DB 494 ThrAlaGlnArgSerLeuSerProLeuHisArgSerIleSerValGluAspAsnIleYr 513
QY 1546 CACACACAGCTTCTCTTTCGCTTTCACACAGCAGCAGACCACTCAAGAGTGTGCTGC 1605
DB 514 HisThrSerPheLeuPheGlyLeuSerThrSerGlnIleHisLeuThrIleSerAlaGly 533

```

```

QY 1606 CTGGGCTTAAAGGCTGGGCACTCGATATCTTGAGCCCCCAGACCTTAACCTTCCTG 1665
DB 534 LeuGlyLeuysrAGlyTyrPheIleAspIleLeuAlaProGlnThrSerThrProSerLeu 553
QY 1666 ACCAGCAGCTGTATTTTCCACAGAGTCTCACTTCTACTCTGCTCAGCCATCTAC 1725
DB 554 ThrSerSerTriPtyrPheAlaThrGluSerSerHisPheTyrSerAlaSerAlaIleYr 573
QY 1726 GAGAGCAGTGGCAGTTACTCTGCTTACAGCTGACAGCCAGCTGCCCCCTTGGAGACCA 1785
DB 574 GlyGlySerAlaSerIleSerIleAlaTyrSerCySerGlnLeuProThrCyseGlyAspGln 593
QY 1786 GTCTATTCTGTGGCAGGCGGAGAGCAAGTGAAGACAGAGCTGACTCGGCGCGAGCTGG 1845
DB 594 ValTyrSerValArgArgArgGlnIleProSerAspArgAlaAspSerArgAspSerTrp 613
QY 1846 CATGAAGAGAGCCCTTTGAAAAAGCAGTTTAAACGAGAAAGCTGCCAAATGAAATTGGA 1905
DB 614 HisGluGluSerProPheGluysrGlnPheIleysrArgArgSerCyseGlnMetGluPheGly 633
QY 1906 GAGACATCATGTCAAGAACAGGTCAAGGAGAGCTGGGAAAGTGGCAGTCACTCT 1965
DB 634 GluSerIleMetSerGluAsnArgSerArgGluGluLeuGlyValGlySerGlnSer 653
QY 1966 AGCTTTGGGCAAGCATGGAATCATTTAGGCTCC 2001
DB 654 SerPheSerGlySerMetGluIleIleGluValSer 665

RESULT 14
AAU79162
ID AAU79162 strand: protein; 665 AA.
AC AAU79162;
XX DT
XX 02-JUL-2002 (first entry)
XX DE
XX Human dual-specificity phosphatase-3 (DSP-16) mutant protein #2.
XX KW
XX Human; dual-specificity phosphatase-3; DSP-16; autoimmune disease;
XX mltogen-activated protein kinase; MAP; Duchenne muscular dystrophy;
XX cancer; graft-versus-host disease; allergy; metabolic disease;
XX abnormal cell growth; abnormal cell proliferation; contact inhibition;
XX cell cycle abnormality; anchorage independent cell growth; apoptosis;
XX intercellular adhesion; DSP-16 modulator; mutant; nuclein.
XX OS
XX Homo sapiens.
XX Synthetic.
XX FH
XX Key Location/Qualifiers
XX FT MISC-difference 244 /note="Wild-type Cys substituted by Ser"
XX FT
XX PN
XX MO200226997-A2.
XX PD
XX 04-APR-2002.
XX XX
XX 25-SEP-2001; 2001WO-US030124.
XX PF
XX 26-SEP-2000; 2000US-0235487P.
XX PR
XX (CEPT-) CEPTYR INC.
XX PA
XX Luche RM, Wei B,
XX PI
XX
XX DR
XX WPI; 2002-315802/35.
XX XX
XX New DSP-16 polypeptide, useful for identifying modulators of its
XX activity, which can be used in the treatment of disorders such as
XX Duchenne muscular dystrophy, or cancer.
XX PT
XX Claim 46; Page; 87pp; English.
XX PS
XX

```

CC The present invention relates to a new polypeptide, DSP-16, having a 665
 CC amino acid sequence, given in the specification, or a variant having at
 CC least 50 % identical residues, which retains the ability to
 CC dephosphorylate an activated mitogen-activated protein (MAP) kinase. The
 CC invention can be used for identifying agents which modulate DSP-16
 CC activity, for modulation of a proliferative response in a cell, survival
 CC of a cell, or differentiation of a cell. The cell displays contact
 CC inhibition of cell growth or anchorage independent growth and may display
 CC altered intercellular adhesion. The agent may modulate apoptosis, or the
 CC cell cycle. The identified modulators can be used to treat Duchenne
 CC muscular dystrophy, cancer, graft-versus-host disease, autoimmune
 CC diseases, allergies, metabolic diseases, abnormal cell growth, abnormal
 CC cell proliferation, and cell cycle abnormalities. The present amino acid
 CC sequence represents the human dual-specificity phosphatase-3 (DSP-16)
 CC mutant protein #2. Note: This sequence is not shown in the specification
 CC but is derived from the wild-type human DSP-16 (AAU79156) protein given
 CC in figure 2 of the specification

XX
 SO Sequence 665 AA:

Alignment Scores:

Pred. No.: 0 Length: 665
 Score: 421.00 Matches: 421
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 63.12% Indels: 0
 DB: 5 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x AAU79162 (1-665)

QY 739 TTATGAGGATCCCGGCTCCGACCATGCTATGCGCTACATGAAAGAGATGAC 798
 Db 245 LENAIGLYLSEETRGSERALATMILEALILELATYRILECTYSAIGECAP 264
 QY 799 ATGCTCTTAGATGAAGCTTACAGATTGTGAAAGAAAAAGACCTACTATCTCCAAAC 858
 Db 265 MESELEUASPGLULATYRARGPHEVALLYSGULYSRPROTHRILESERPROAN 284
 QY 859 TTCAATTTTGGGCCAATCTCTGACTATGAGAAAGATTAAAGACGACTGGAGCA 918
 Db 285 PHEASNPHELEUGLYNLEULEUASPTRYGLULYELYSILEYSSANGINTHRYALAA 304
 QY 919 TCAGGGGCAAAAGCAAACTCAAGCTGCTGACCTGAGAAAGCAATGAACCTGTCCCT 978
 Db 305 SERGLYPROLYSERLYSEULYSLEULEHLSLEUGLULYSRPROANGULPROVALPRO 324
 QY 979 GCTGTCTCAGAGGGTGAGACAGAAAGGAGACGCCCTCAGTCACCTGTGCGACTCT 1038
 Db 325 ALAVALSERGLUGLYGLINLYSSERGLUTHRPROLEUSERPROCYSAIAAPSER 344
 QY 1039 GCTAAGCTCAGAGGACAGACAGCAAAAGCCGCTGATCCGCCAGCGTCCAGCTGCC 1098
 Db 345 ALATHSERGLULALAGLYGINARGPROVALHAPROVALSERVALPROSERVALPRO 364
 QY 1099 AGCGTGAGCGCTCGCTGTAGAGAGACAGCCCGCTGTACAGCGCTCAGTGGCTGCAC 1158
 Db 365 SERVALGINPROSERLEULEUENGLUASERPEROLEUVALGINALALEUSERGLYLEUHS 384
 QY 1159 CTGTCCGACAGACAGCTGGAAGAAGACGCAATTAAGCTCAAGGCTTCTTCTGTGATATC 1218
 Db 385 LEUSERALASPARGLEUGLULASERASNLYSLEULYSRGERPHESELEUASPILE 404
 QY 1219 AAATCAGTTTCATATTCAGGACGATGCGACATCTCTTACATGGCTTCTCTCATAGAA 1278
 Db 405 LYSERVALSERLYSERALASERMETVALALASERLEUHSGLYPHESESESERSEGLIN 424
 QY 1279 GATGCTTTGGAATACATCAAACTTCACACTCTGGATGGAGCAACAGACTATGCGAG 1338
 Db 425 ASPHALALEULIULYRILYRILYSPROSETRINTHLEUASPGLYTHINLEULYSCSGLIN 444
 QY 1339 TTCTCCCTGTTCAGAACTATTCGAGACGACTCCGAAACCAAGTCTGTATTAAGAGAA 1398
 Db 445 PHESEPRVALGINGLINLEUSERGLUGINTHRPROGLINTHRSEPRVALYSGLUGLIN 464

QY 1399 GCCACATCCCAAGAACTGCAAGCCGACGCTTACAGACAGCAGAGGATTTG 1458
 Db 465 ALASERILEPROLYSLEUGLINTHRLALARGPROSEAPSERGLINSELYARGLEU 484
 QY 1459 CATTTGCTCAGAACAGAGACAGATGGACCCGCCAGAGGTCCTTTATCTCACTGCAT 1518
 Db 485 HISSEVALARGTHRSESESERGLYTHRALAGINASERLEUSERPROLEUHS 504
 QY 1519 CGAAGTGGAGCGTGGAGACAAATTAACAACACAGCTTCTTTGGCCTTTCCACGAC 1578
 Db 505 ARGSEGLYSERVALGLUASPARNTYRHSRHSERPHELEUPHEGLYLEUSERTHSR 524
 QY 1579 CAGCAGACCTCAGAAAGTCTGGCGCTGGGCTTTAGAGGCTGACATCGGATATCTTG 1638
 Db 525 GINGLINHSLEUTHNLYSERALAEGLYLEUGLYLEULYSGLYTPHISERASPILEU 544
 QY 1639 GCCCCCAAGACTTACCCCTTCCCTGACACAGACCTGTATTTGCCACAGATCTCA 1698
 Db 545 ALAPROGLINTHRSETRHROSERLEUTHRSESETRPYRPHALATHRGLUSER 564
 QY 1699 CACTTCTACTGTGCTCAGCAATCTACGAGGACGTCAGATTACTTGCTTACAGCTGC 1758
 Db 565 HSPHERYSEVALASERALAEGLYGLYSEVALASERTYRSEVALATYSECY 584
 QY 1759 AGCCAGCTGCCCACTTTGGGAGACCAAGCTTATCTGCGCAGGCGGACAGCAAGT 1818
 Db 585 SERGLINLEUPROTHNLYSGILYASPGINALLYRSEVALARGARGARGINLYSEPRO 604
 QY 1819 GACAGAGCTGACTCGCGCGAGCTGCGATGAGAGAGACCCCTTTGAAAGCAGTTTAA 1878
 Db 605 ASPARGALASPSEARARGSERTRPHISGLUGLUSERPROPHGLULGLINPHELYS 624
 QY 1879 CGCAAGCTGCGCAATATGGAATTGGAGAGACATCATGTCAAGAAACAGTTCAGGGAA 1938
 Db 625 ARGARGSECYSGINMETGLUPHEGLYUSERILEMETSERGLUASARGSERARGGLU 644
 QY 1939 GAGTGGGGAAAGTGGGAGTCAAGCTGAGCTTTTGGGACAGCATGAATCATTTAGAGT 1998
 Db 645 GLULEUGLYLVALGILYSEGLINSESESEPHESERGLYSEMERGLULILEGLUVAL 664
 QY 1999 TCC 2001
 Db 665 Ser 665

RESULT 15
 ID ABR52352
 ID ABR52352 strand1; protein; 665 AA.
 AC ABR52352;
 DT 19-JUN-2003 (first entry)
 XX
 DE Protein relating to the invention SEQ ID NO: 42.
 KW antiproliferative; hepatotropic; nephrotropic; antiarthritic;
 KW antipsoriatic; candiant; cytostatic; gene therapy; liver disease;
 KW proliferative disorder; renal failure; cardiovascular disorder;
 KW immunological disorder; arthritis; psoriasis; congenital heart defect;
 KW congestive heart failure; cancer; human; mouse; yeast; rat; fruitfly.
 OS Homo sapiens.
 PN WO200257460-A2.
 PD 25-JUL-2002.
 PF 20-DEC-2001; 2001WO-US050459.
 PR 20-DEC-2000; 2000US-0256868P.
 PR 30-MAR-2001; 2001US-0280186P.
 PR 01-MAY-2001; 2001US-0287735P.
 PR 05-JUN-2001; 2001US-0295848P.

25-JUN-2001; 2001US-0300465P.

(BRIM) BRISTOL-MYERS SQUIBB CO.

Jackson DG, Feder J, Nelson T, Mintier G, Ramanathan C, Lee L, Slemers N, Bol D, Schieven G, Finger J, Toddered CG, Bassolino D, Krytek S, Mcatee P, Suchard S, Banas D,

WPI: 2002-599721/64.

N-PsDB; ACC60521.

Novel polynucleotides encoding human phosphatase polypeptides useful in the prevention or treatment of e.g. proliferative and cardiovascular disorders.

Claim 5, Fig 5, 801pp; English.

The invention relates to a novel isolated nucleic acid comprising a polynucleotide having a nucleotide sequence selected from 40 polynucleotides fully defined in the specification. The polynucleotide of the invention has antiproliferative, hepatotropic, nephrotropic, antiarthritic, antiprostatic, cardiac, and cytostatic activity. The polynucleotide may have a use in gene therapy. A polynucleotide or polypeptide of the invention is useful for preventing, treating or ameliorating a medical condition, e.g. a proliferative disorder. They are also useful for treating e.g. liver disease, renal failure, immunological disorders including arthritis and psoriasis, cardiovascular disorders such as congenital heart defects and congestive heart failure, and cancer. A method of the invention is useful for diagnosing a pathological condition or susceptibility to a condition in a subject. The present sequence is used in the exemplification of the invention

Sequence 665 AA;

Alignment Scores:	
Pred. No.:	0
Score:	350.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	52.47%
DB:	5
	Length:
	Matches: 565
	Conservative: 350
	Mismatches: 0
	Indels: 0
	Gaps: 0

QY	952	CTGGAGGAGCCAAATGAACCTGCTCCCTGCTGCTCAGAGCGGTGACACAAAAAGCCAGAC	1011
Db	316	LeuGluIuVProAsnGluProValProAlaValSerGluGluGlyGlnIlySerGluThr	335
QY	1012	CCCCCAGTCCACCCCTGTGGCCGACCTGTGCTACTCAGAGGCGACGAGCAAAAGGCCGTG	1071
Db	336	ProLeuSerProProCysAlaAspSerAlaThrSerGluAlaAlaGlyGlnIrrProVal	355
QY	1072	CATCCCGGACGCGTGCACGCGTGCACGCGTGCACGCGTGCCTGTGTAGAGGACGACCGCG	1137
Db	336	HisProAlaSerValProSerValProSerValGlnProSerLeuLeuGluAspSerPro	375
QY	1132	CTGTATACAGAGCCTCAGTGGCGGTGACCCGTGCCGACAGCAGCCTGAAAGACAGCAATAG	1191
Db	376	LeuValGlnAlaLeuSerGlyLeuAlaIleuSerAlaAspArgLeuGluAspSerAsnIly	395
QY	1192	CTCAAGCCTTCTCTCTCTGAGTATCAAAATAGTTTCAATATTCAGCCAGCAGTGGCAGCA	1251
Db	396	LeuIlyValGlySerPheSerLeuAspIleIlyAspValSerTyrSerAlaSerMetAlaIa	415
QY	1252	TCCCTTACATGGCCTCTCTCATCAGAAAGATGCTTTGGATTACTACAAACCTTCCACTACT	1311
Db	416	SerLeuHisGlyPheSerSerSerGluAspAlaLeuGluIlyTyrIlyPheSerThrThr	435
QY	1312	CTGGATGGGACCCAAACGCTATGCGAGTTCTCCCTGTTCAGAACTATCGGACGAGACT	1371
Db	436	LeuAspGlyThrAsnIlyLeuIlyCysGlnPheSerProValGlnGluLeuSerGluGlnThr	455
QY	1372	CCCGAAACCACTCTGATTAAGAGGAAAGCCAGCATCCCCAGAAAGCTGCAGACGCCCAGG	1431

Db	456	ProtilinhserrroabrylsglignlalserrlerolybybyleuGlntrhAlaArg	475
Oy	1432	CTTTGACAGACCCAGACCAAGCATTTGCATTCCGTCGAAGAACAGACAGTGGACCCG	1491
Db	476	ProberhserserglnserlybArgluehlnlservalArgthrserSerGlyThrAla	495
Oy	1492	CAGAGGTCCTTTTATTCCTACATGCATCGAAGTGGAGGCGCTGGAGGACATTTACACAC	1551
Db	496	GlntrgserrleuenserProleuhlnsArgserGlyserValGluapaaenlyrhsthr	515
Oy	1552	AGCTTCCTTTTCGGCCTTTCCACAGCCAGCAGCAGCAGCAGTCTGCTGGCTGGG	1611
Db	516	SertheuhsrpeglyleuserThrserGlnGlnhlsheuthrlysserrAlaGlyleuGly	535
Oy	1612	CTTAAGGCTGGGCACTCGAATATCTTGCCCCCGACACTTACCCCTTCCTGACACG	1671
Db	536	LeuylsglytrphlserrapllleleuAlaproGlnthrserThrProserleuthrser	555
Oy	1672	AGCTGGATTTTGGCCACAGAGTCTCTCACACTTCTACTCTGCTGCCTTCAGCATTCAGAGGC	1731
Db	556	SerrtrpyrPhealatrngluserSerhlsphetyrSerAlaserrAlaIetyrGlyGly	575
Oy	1732	AGTGCACGTTACTGTGCTACAGCTGCAGCAGCTGCCACTTGGCGGAGCAAGTCTAT	1791
Db	576	SerrlaserlyrserAlaIytrserCysserGlnneurOthrCybglyaberglnValtyr	595
Oy	1792	TCTGTGGCAGAGCGCGCAAGCCAAAGTACAGAGCTACTCGCGCGAGCTGGCATGAA	1851
Db	596	SerrAlaItrghArglnlyrProserbArghAlaerperhArgAserserrTrpHleGlu	615
Oy	1852	GAGAGCCCTTTGAAAAGCAGTTTAAAGCAAGAGCTGCCAAATGCAATTTGGAGAGAC	1911
Db	616	GlnserProppreglulysGlnphelysArgArgserCysslnnetGlnPregllyGlnser	635
Oy	1912	ATCATGTACAGAGACAGGTACACGGGAGAGACTGGGGAGAGGGGAGTCAAGTACGTTT	1971
Db	636	IlelterserGluanaArgserArgGlnGlnuendyluValGlyserGlnserSerphe	655
Oy	1972	TCGGGAGCATGGAATCATTTAGAGGTTCC	2001
Db	656	SerGlyserMetGlnIleIleGlnValSer	665
RESULT 16			
AAB20325			
ID	AAB20325 standard; protein; 666 AA.		
AC	AAB20325;		
XX			
DT	29-MAY-2001 (first entry)		
XX			
DE	Human protein phosphatase and kinase protein-4.		
XX			
KM	Protein phosphatase and kinase protein; PPKP-4; human; gaetrofretstinal disorder; immune system disorder; neurological disorder; cell proliferative disorder; cancer; diagnosis; therapy.		
KM			
XX			
OS	Homo sapiens.		
XX			
FT	Key	Location/Qualifiers	
FT	Modified-site	12	/note= "O-phosphorylated"
FT	Modified-site	21	/note= "O-phosphorylated"
FT	Modified-site	23	/note= "O-phosphorylated"
FT	Modified-site	38	/note= "O-phosphorylated"
FT	Modified-site	38	/note= "N-glycosylated"
FT	Modified-site	38	/note= "N-glycosylated"
FT	Modified-site	49	/note= "N-glycosylated"
FT	Modified-site	72	/note= "N-glycosylated"

PT	Modified-site	/note= "O-phosphorylated"
FT		82
FT		/note= "O-phosphorylated"
FT	Region	85. .298
FT		/note= "VH1-type dual specificity phosphatase signature"
FT	Modified-site	91
PT		/note= "O-phosphorylated"
FT	Modified-site	190
FT		/note= "N-glycosylated"
FT	Modified-site	212
FT		/note= "N-glycosylated"
FT	Modified-site	214
FT		/note= "O-phosphorylated"
FT	Active-site	220. .280
FT		/note= "tyrosine specific protein phosphatase"
FT	Region	237. .278
FT		/note= "y phosphatase signature"
FT	Modified-site	266
FT		/note= "O-phosphorylated"
FT	Modified-site	280
FT		/note= "O-phosphorylated"
FT	Modified-site	300
FT		/note= "N-glycosylated"
FT	Modified-site	369
FT		/note= "O-phosphorylated"
FT	Modified-site	393
FT		/note= "O-phosphorylated"
FT	Modified-site	421
FT		/note= "O-phosphorylated"
FT	Modified-site	422
FT		/note= "O-phosphorylated"
FT	Modified-site	434
FT		/note= "O-phosphorylated"
FT	Modified-site	439
FT		/note= "O-phosphorylated"
FT	Modified-site	468
FT		/note= "O-phosphorylated"
FT	Modified-site	471
FT		/note= "O-phosphorylated"
FT	Modified-site	479
FT		/note= "O-phosphorylated"
FT	Modified-site	528
FT		/note= "O-phosphorylated"
FT	Modified-site	590
FT		/note= "O-phosphorylated"
FT	Modified-site	597
FT		/note= "O-phosphorylated"
FT	Modified-site	605
FT		/note= "O-phosphorylated"
FT	Modified-site	610
FT		/note= "O-phosphorylated"
FT	Modified-site	613
FT		/note= "O-phosphorylated"
PT	Modified-site	618
FT		/note= "O-phosphorylated"
FT	Modified-site	628
FT		/note= "O-phosphorylated"
FT	Modified-site	641
FT		/note= "N-glycosylated"
PT	Modified-site	643
FT		/note= "O-phosphorylated"
XX		WO200120004-A2.
XX		22-MAR-2001.
XX		14-SEP-2000; 2000WO-US025515.
XX		15-SEP-1999; 99US-0154141P.
XX	(INCY-) INCYTE GENOMICS INC.	
PI	Yue H, Tang YT, Bandman O, Hillman JL, Baughn MR, Azimzai Y,	

PI	Lv DAM;
XX	
DR	WPI; 2001-244811/25.
XX	N-PsDB; AAF30479.
PT	Novel human protein phosphatase and kinase proteins for diagnosis,
PT	treatment and prevention of gastrointestinal, immune system, neurological
PS	and cell proliferative disorders.
XX	
PS	Claim 1; Page 87-88; 103pp; English.
CC	The present sequence is that of novel human protein phosphatase and
CC	kinase protein PPHK-4, as predicted from Incyte Clone ID No. 1234795CB1
CC	(see AAF30479). Tissues that express PPHK-4 (as a fraction of total
CC	tissue) expressing PPHK-4 include gastrointestinal (0.385),
CC	cardiovascular (0.154), nervous (0.154) and reproductive (0.154).
CC	Diseases or conditions associated with tissues expressing PPHK-4 (as a
CC	fraction of total tissues expressing PPHK-4) include cancer (0.692),
CC	inflammation or trauma (0.308) and cell proliferation (0.231). The
CC	encoded protein shows homology to mouse neuronal tyrosine threonine
CC	phosphatase 1. The invention provides human PPHK-1 to -11 polypeptides
CC	(see AAB20322-32) and polynucleotides (see AAF30476-86). It also provides
CC	expression vectors, host cells, antibodies, agonists and antagonists, as
CC	well as methods for diagnosing, treating or preventing disorders
CC	associated with expression of PPHK, including gastrointestinal
CC	disorders, immune system disorder, neurological disorders and cell
XX	proliferative disorders, including cancer
SQ	
Sequence 666 AA;	
Alignment Scores:	
Pred. No.:	2,97e-303 Length: 666
Score:	318..00 Matches: 471
Percent Similarity:	99.58% Conservative: 0
Best Local Similarity:	99.58% Mismatches: 2
Query Match:	47.68% Indels: 2
Gaps:	4 Gaps: 0
US-10-029-345A-108_COPY_532_2532 (1-2001) x AAB20325 (1-666)	
OY	ACCGTCGCAAAAGCCTACTTATCCGCAGTGTCATTTCGCCGGGTGCTGGATGACC 645
Db	
194	TnrcPSPolySProAsPhnelProGlnSerHisPhneLysValProAlaInsbnp 213
OY	646 AGCTTTTGAGAATAATTTCGCCGTGTGGACAATAGTAGTGAATTCATTGAGAAGCA 705
Db	
214	SerPiecSGIuLySIleLeuProtRipneuabpySSerValAspHneiLeiulySaia 233
OY	706 AAAGCTCATGANTGATGTTCCTTAAGTCACTGTTAGCTGGATCTCCCGCTCCGCACC 765
Db	
234	LysAlaSeraAnglyCyvalLeuValHisCylSleuAdaglyllLSerArGerlaThr 253
OY	766 ATCGGTATCGGCTACATCATGAAGAAGATGGACATGTCTTTAGATGAAGCTTACAGATT 825
Db	
254	IleaAtIleaIaTyrlImetElysArgMetSerLeuASpGIuaIaTyrrgPre 273
OY	826 GTGAAGAAGAAAAGCCTACTATATCTCCAACCTCAATTTTCTGGGCCAACCTCTGGAC 885
Db	
274	VallysgIuLySBargProtoThriLeserProAnPrheaBpnheuleuclyGIInleueunap 293
OY	886 TATGAGAAAGAGATTAGAACCCAGACTGGAGCATCGAGGCCCAAAGACAACTCAAGCTG 945
Db	
294	TyrGIuLySluSIleLeuAbnglinThrISyLaSerGIuPolySerILeuLySleu 313
OY	946 CTGCACCTGAGAAAGCCAAATGAACCTGTCTCTGTCCTAGAGAGGTGGACAGAAAAAGC 1007
Db	
314	LeuhIsleugIuLyPSroAnsgluProVaIProAlaValISergluNglyGIInlysser 333
OY	1006 GAGAGCGCCCTCAGGCCAAGCCTGGGCCGACTCTGTGTACCTCAGAGAGGACAGAAAGG 1066
Db	
334	GlumrProleuSerProProcycyHaIAspseralatrSerGIuaIaIagiIuinrg 353
OY	1066 CCGGTGATCCCGCAGCGCTGCCAGCTGGCCCAGCGTGCAGCCCTGCTGTTAGAGNAC 1122


```

DB      354  ProValHisProIaSerValIProSerValIProSerValGlnProSerLeuLeuGlnuAAsp 373
QY      1126  AGCCCGCTGGTACAGGCGCTCACTGGAGCTGCACCTGTCCGACAGACAGCTGGAGAAGCAGC 1185
DB      374  SerProLeuValGlnIaLeuSerGlyLeuHisLeuSerAlaAspArgLeuGlnuAAspSer 393
QY      1186  AATTAAGTCAAGGCTTCTTCTCTGAGATATCAAAATCAATTCAATTTACAGCAGATG 1245
DB      394  AsnLysLeuLysArgSerPheSerLeuAspIleLysSerValSerLysSerAlaSerMet 413
QY      1246  GCAGACCTCTTAATGAGCTTCTCTCATGAGAAAGATGCTTGGAAATATCAAAACCTTCC 1305
DB      414  AlaIleSerLeuHisGlyPheSerSerSerGlnuAspAlaLeuLysLysArgProSer 433
QY      1306  ACTACTCTGGATGGAGCAACAAGCTATGTCAGCTTCTCCCTGTGACAGAACTATCGAG 1365
DB      434  ThrThrLeuAspArgLysThrAsnLysLeuGlnPheSerProValGlnGlnuLeuSerGln 453
QY      1366  CAGACTCCCGAAACCAAGTC--CTGATTAAGAGAGAGACAGACATCCCAAGAGCTGCAGA 1423
DB      454  GlnThrProGlnLysSerSerLeuIleArgArgLysProIaSerProArgSerCysArg 473
QY      1424  CGCGCAGGCTTACAGACAGCAGAGAGGATTTGCTGGTCAGAGAACAGAGAGAGT 1483
DB      474  ProGlnLysLeuGlnThrAlaArgAlaSerAspCysIleArgSerGlnuProIaAlaVal 493
QY      1484  GCACCGCCCGAGAGCTCCCTTTATCTCACTGATGCAATGAGAGAGCTGGAGAGAAAT 1543
DB      494  AlaProProArgGlyProPheGlyLysLeuHisCysIleGlnValGlyLysAlaArgThrIle 513
QY      1544  ACCACACAGCTTCTTCTTGGCCCTTTCACACAGCCAGACAGACCTTCAGAGAGCTGCTG 1603
DB      514  ThrThrProIaSerPheSerAlaPheProProIaSerSerThrSerArgSerLeuLeu 533
QY      1604  GCCTGGGC--CTTAAGGCTGGGCACTGGGATATCTGGCCCCCGACGCTTACCCCTTCC 1662
DB      534  AlaTrpAlaLeuLysGlyTrpHisSerAspIleLeuAlaProGlnThrSerTrpProSer 553
QY      1663  CTGACCAGCAGCTGTGATTTTGGCACAGAGTCTTCACACTTACTCTGCTCAGCCATC 1722
DB      554  LeuThrSerSerTrpTrpPheAlaThrGlnSerSerHisPheLysSerAlaSerAlaIle 573
QY      1723  TACGAGAGCAGTGCAGTTACTTGTCCCTACAGCTGCAGCAGCTGCCACTTCCGGAGAC 1782
DB      574  TyrGlyGlySerAlaSerLysSerAlaLysSerCysSerGlnuLeuProThrCysGlyAsp 593
QY      1783  CAAGTCAATCTGTGGGCAAGCGGCAAGACCAAGTGAAGAGAGCTGACTGGCGGGGAGC 1842
DB      594  GlnValLysSerValAlaArgArgGlnLysProSerTrpArgAlaAspSerAlaArgSer 613
QY      1843  TGGCATGAAGAGAGCCCTTTGAAAAGCAAGTTTAAAGCAGAAGCTGCCAATGAATTT 1902
DB      614  TrpHisGlnGlnuSerProPheGlnuLysGlnPheLysArgArgSerCysGlnuMetGlnuPhe 633
QY      1903  GGAAGAGCATCATGTGACAGAAACAGGCTCAGGAAAGAGCTGGGAGAAAGTGGCAGTCA 1962
DB      634  GlyLysSerLysLeuMetSerGlnuAsnArgSerArgGlnuLysValGlySerGln 653
QY      1963  TCTAGCTTTTGGGAGCAGATGAATCATGAGAGTCTCC 2001
DB      654  SerSerPheSerGlySerMetGlnuIleGlnuLysSer 666

```

RESULT 17

ABR52424 standard; protein; 664 AA.

AC ABR52424;

DT 19-JUN-2003 (first entry)

XX Protein relating to the invention SEQ ID NO: 190.

XX

KW antiproliferative; hepatotropic; nephrotropic; antiarthritic;
 KW antipneumatic; cardiac; cytosolic; gene therapy; liver disease;
 KW proliferative disorder; renal failure; cardiovascular disorder;
 KW immunological disorder; arthritis; psoriasis; congenital heart defect;
 KW congestive heart failure; cancer; human; mouse; yeast; rat; fruitfly.

OS Homo sapiens.

PN WO200257460-A2.

XX 25-JUL-2002.

XX 20-DEC-2001; 2001WO-US050459.

XX 20-DEC-2000; 2000US-0256868P.

XX 30-MAR-2001; 2001US-0280186P.

XX 01-MAY-2001; 2001US-0287735P.

XX 05-JUN-2001; 2001US-0295848P.

XX 25-JUN-2001; 2001US-0300465P.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.

XX Jackson DG, Feder J, Nelson T, Mintier G, Ramanathan C, Lee L;

XX PI Siemera N, Bol D, Schieven G, Finger J, Todderud CG, Bassolino D;

XX PI Kryszek S, Mcatee P, Suchard S, Banas D;

XX WPI; 2002-599721/64.

XX Example 57, Page 500-501, 801pp; English.

XX The invention relates to a novel isolated nucleic acid comprising a
 CC polynucleotide having a nucleotide sequence selected from 40
 CC polynucleotides fully defined in the specification. The polynucleotide of
 CC the invention has antiproliferative, hepatotropic, nephrotropic,
 CC antiarthritic, antipneumatic, cardiac, and cytosolic activity. The
 CC polynucleotide may have a use in gene therapy. A polynucleotide or
 CC polypeptide of the invention is useful for preventing, treating or
 CC ameliorating a medical condition, e.g. a proliferative disorder. They are
 CC also useful for treating e.g. liver disease, renal failure, immunological
 CC disorders including arthritis and psoriasis, cardiovascular disorders
 CC such as congenital heart defects and congestive heart failure, and
 CC cancer. A method of the invention is useful for diagnosing a pathological
 CC condition or susceptibility to a condition in a subject. The present
 CC sequence is used in the exemplification of the invention

XX Sequence 664 AA;

XX SQ

XX Alignment Scores:

XX Pred. No.:

XX Score:

XX Percent Similarity:

XX Best Local Similarity:

XX Query Match:

XX DB:

XX US-10-029-345a-108_COPY_532_2532 (1-2001) x ABR52424 (1-664)

XX QY 913 GGAAGCATCAGGCGCAAGCAAACTCAAGCTGCTGAGAGAAAGCAATGAACCT 972

XX DB 302 GtYAlaSerGlyProLysSerLysLysLysLeuLysLeuHisLeuGlnuLysProAsnGlnuPro 321

XX QY 973 GTCCCTGCTGTGTCAGAGGCTGGAGCAAGAAAGGAGAGCCCTCAGTCAAGCTGTGCC 1032

XX DB 322 ValProIaValSerLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 341

XX QY 1033 GACTGTGTAATCTCAGAGGCGAGAGGAGCAAGAGCCCGTGCATCCCGCAGCTGCCAG 1092

XX DB 342 AspSerAlaThrSerGlnuAlaGlnuLysGlnuLysGlnuLysGlnuLysGlnuLysGlnu 361

QY 1093 GTGCCAGCCGTGACCCGTCGTTAGAGAGACCCGCTGATACAGCGCTCAGTGGG 1152
|
|
|
Db 362 ValProSerValGlnProSerLeuLeuGlnAspSerProLeuValGlnAlaLeuSerGly 381
|
|
|
QY 1153 CTGACCTGTTCGACGACAGGCTGAAAGACAGCAATTAAGCTCAAGCCTTCTCTCTG 1212
|
|
|
Db 382 LeuHisLeuSerAlaAspArgLeuGlnAspSerAsnGlyAspSerLeuValArgSerPheSerLeu 401
|
|
|
QY 1213 GATATCAATCAGTTTCACTATATTCAGCCAGCATGCGACATCTTCAATNGCTTCTCTCA 1272
|
|
|
Db 402 AspiLeuSerValSerTyrSerAlaSerMetAlaAlaSerLeuHisGlyPheSerSer 421
|
|
|
QY 1273 TCAGAAATGCTTTGGAATACATACTTCAAACTTCCACTCTGATGAGCAACAAGCTA 1332
|
|
|
Db 422 SerGlnAspAlaLeuGlnTyrTyrTyrLeuProSerThrThrLeuAspGlyThrAsnGlyLeu 441
|
|
|
QY 1333 TGCACATTTCCCTGTTTTCAGAACTATCGAGAGAGACTCCGAAACAGTCTGATTAAG 1392
|
|
|
Db 442 CysGlnPheSerProValGlnGlnLeuSerGlnGlnThrProGlnThrSerProAspGlyS 461
|
|
|
QY 1393 GAGGAAGCCAGCATCCCAAGAAAGCTGACAGCCGCGGCTTACAGACAGCCAGAGCAG 1452
|
|
|
Db 462 GlnGlnAlaSerLeuProValGlnGlnLeuSerGlnGlnThrAlaArgProSerAspSerGlnSerLeuS 481
|
|
|
QY 1453 CGATTGCATTGGTTCAGAACCCAGCAGCAGTGGCACCCGCAAGTCCCTTTATCTCCA 1512
|
|
|
Db 482 ArgLeuHisSerValArgThrSerSerSerGlyThrAlaGlnArgSerLeuLeuSerPro 501
|
|
|
QY 1513 CTGATGATGAGTGGGACCGTGGAGACAAATTACCAACCCAGCTTCTTTCGCGCTTTTC 1572
|
|
|
Db 502 LeuHisArgSerGlySerValGlnAspAsnTyrHisThrSerPheLeuPheGlyLeuSer 521
|
|
|
QY 1573 ACCAGCCAGCAGCAGCACTCAGAAAGTGTGCTGCGCTGGGCTTAAAGGGCTGGCACTCGAGAT 1632
|
|
|
Db 522 ThrSerGlnGlnHisLeuThrLysSerAlaGlyLeuGlyLeuLysGlyTyrPheAspSer 541
|
|
|
QY 1633 ATCTTGACCCCAAGACCTTACCCCTTCCCTTACCAAGCAGCTGTATTTTGGCACAGAG 1692
|
|
|
Db 542 IleLeuAlaProGlnThrSerThrProSerLeuThrSerSerTyrPheAlaThrGln 561
|
|
|
QY 1693 TCCCTCACACTTACTCTGCTCAGCAGCATTAAGGAGGAGCGCCAGTTACTCTGCTTAC 1752
|
|
|
Db 562 SerSerHisPheLysSerAlaSerAlaIleTyrGlySerAlaSerTyrSerAlaTyr 581
|
|
|
QY 1753 AGC 1755
|
|
|
Db 582 Ser 582
|
|
|
RESULT 18
AAM25744
ID AAM25744 standard; protein; 672 AA.
AC AAM25744;
XX
XX
DT 16-OCT-2001 (first entry)
XX
XX
DE Human protein sequence SEQ ID NO:1259.
XX
XX
KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
KW antiinflammatory; antirheumatic; antithyroid; immunosuppressive;
KW antibacterial; endocrine; cardiac; central nervous system; vitruclide;
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
KW antiaggregant; haemostatic; vulnery; antitumor; osteoporosis; eczema;
KW dermatological; antiallergic; antisthmatic; antidiabetic; cytostatic;
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
KW immunosuppressant; gene therapy; antitense therapy; vaccine; inflammation;
KW immunophylactic; rheumatoid arthritis; septic shock; pancreatitis;
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
KW allergic rhinitis; diabetes; multiple sclerosis; depression;
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
KW neurological disorder.

XX
OS Homo sapiens.
XX
EN MO200153455-A2.
XX
PD 26-JUL-2001.
XX
PF 22-DEC-2000; 2000MO-US035017.
XX
PR 23-DEC-1999; 99US-00471275.
PR 21-JAN-2000; 2000US-00488725.
PR 25-APR-2000; 2000US-00552317.
XX
PA (HYSR-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR MPI; 2001-457603/49.
XX
DR N-PSDB; AAH99685.
XX
PT Isolated human polynucleotides encoding polypeptides, useful for the
XX
XX treatment and diagnosis of e.g. cancer, ulcers and HIV infection.
XX
PS Claim 20; Page 260; 1217p; English.
XX
CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
CC AAM25963. The proteins can have activities based on the tissues and cells
CC they are expressed in, such as: antiinflammatory; antirheumatic;
CC antiallergic; immunosuppressive; antibacterial; endocrine; cardiac;
CC central nervous system; vitruclide; anti-HIV; fungicide; antimutagen;
CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnery;
CC antitumor; osteopathic; dermatological; antiallergic; antisthmatic;
CC antidiabetic; cyrostatic; neuroprotective; antidepressant; nootropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antitense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders
XX
SQ Sequence 672 AA;
XX
Alignment Scores:
Pred. No.: 3,02e-244 Length: 672
Score: 258.00 Matches: 258
Percent Similarity: 100.00% Conservativity: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 38.68% Indels: 0
DB: 4 Gaps: 0
US-10-029-345a-108_COPY_532_2532 (1-2001) x AAM25744 (1-672)
QY 586 ACCTGTCCAAAGCCTGACCTTATCCCGAGTCTGATTTCTGCGTGGCTGGAATGAC 645
|
|
|
Db 201 ThrCysProLysProAspPheIleProGlnSerHisPheLeuArgValProValAsnAsp 220
|
|
|
QY 646 AGCTTTGTGAGAAATTTTGGCCGTGTGACAAATCAGTAATTTTCAATTGGAAGAACA 705
|
|
|
Db 221 SerPheCysGlnLysIleLeuProTyrPheAspLysSerValAspPheIleGlnVala 240
|
|
|
QY 706 AAAGCTCCAAATGATGTGTTCTAGTCACTGTTTACGTGGATTCGCGCTCGCACACC 765
|
|
|
Db 241 LysAlaSerIleGlnGlyValLeuValHisCysLeuAlaIleGlyIleSerArgSerAlaThr 260
|
|
|
QY 766 ATGCGTATGCGCTTACATCATGAAGAGATGACATGCTTTAGATGAAGCTTACAGATT 825
|
|
|

XX 10-DEC-2001; 2001US-0339739P.
PR 11-DEC-2001; 2001US-0339453P.
PR 14-MAR-2002; 2002US-0365091P.
PR 14-MAR-2002; 2002US-0365384P.
PR 12-APR-2002; 2002US-0372381P.
PR 12-APR-2002; 2002US-0372615P.
PR 22-APR-2002; 2002US-00128558.
PR 24-APR-2002; 2002US-0376045P.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Asundi V, Goodrich RW, Ren P, Zhang J, Zhao QA, Wang J;
PI Ghosh M, Xue AJ, Wehrman T, Meng G, Zhou P, Drmanac RT, Wang Z;
PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ,
XX
DR WPI: 2003-569235/53.
DR N-PSDB; ADE07547.
XX
PT New polynucleotides, useful for expressing recombinant proteins for
PT analysis, characterization or therapeutic use, or as markers for tissues
PT in which the corresponding protein is preferentially expressed.
XX
PS Claim 20; SEQ ID NO 1524; 1177bp; English.
XX
CC The invention comprises the amino acid and coding sequences of novel
CC proteins. The DNA and protein sequences of the invention are useful as:
CC markers for tissues in which the corresponding protein is preferentially
CC expressed; as molecular weight markers on gels; as chromosome markers or
CC tags; to identify chromosomes or to map related gene positions; and to
CC compare with endogenous DNA sequences in patients to identify potential
CC genetic disorders. The present amino acid sequence represents a protein
CC of the invention.
XX
SQ Sequence 579 AA;
XX
Alignment Scores:
Pred. No.: 3.57e-10 Length: 579
Score: 20.00 Matches: 20
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.00% Indels: 0
DB: 7 Gaps: 0
US-10-029-345A-108_COPY_532_2532 (1-2001) x ADE08458 (1-579)
OY 730 GTGCAGTGTGATGCGGATCTCCGCTCCGACCATGCGCTATCGCTACATCATGANG 789
Db 186 ValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIleMetLys 205
RESULT 21
ABR52382
ID ABR52382 standard; protein; 625 AA.
XX
AC ABR52382;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein relating to the invention SEQ ID NO: 110.
XX
KW anti-proliferative; hepatotropic; nephrotropic; antiarthritic;
KW anti-psoriatic; cardiant; cytosatic; gene therapy; liver disease;
KW proliferative disorder; renal failure; cardiovascular disorder;
KW immunological disorder; arthritis; psoriasis; congenital heart defect;
KW congestive heart failure; cancer; human; mouse; yeast; rat; fruitfly.
OS Homo sapiens.
XX
PN WO200257460-A2.
XX
PD 25-JUL-2002.
XX
PF 20-DEC-2001; 2001WO-US050459.

XX 20-DEC-2000; 2000US-0256868P.
PR 30-MAR-2001; 2001US-0280186P.
PR 01-MAY-2001; 2001US-0287735P.
PR 05-JUN-2001; 2001US-0295848P.
PR 25-JUN-2001; 2001US-0300465P.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Jackson DG, Feder J, Nelson T, Mintier G, Ramanathan C, Lee L;
PI Stiemens N, Bol D, Schieven G, Finger J, Todderud CG, Baasolino D;
PI Kysceck S, Mcatee P, Suchard S, Banas D;
XX
DR WPI: 2002-599721/64.
XX
PT Novel polynucleotides encoding human phosphatase polypeptides useful in
PT the prevention or treatment of e.g. proliferative and cardiovascular
PT disorders.
XX
PS Disclosure; Fig 14; 801bp; English.
XX
CC The invention relates to a novel isolated nucleic acid comprising a
CC polynucleotide having a nucleotide sequence selected from 40
CC polynucleotides fully defined in the specification. The polynucleotide of
CC the invention has anti-proliferative, hepatotropic, nephrotropic,
CC antiarthritic, anti-psoriatic, cardiant, and cytosatic activity. The
CC polynucleotide may have a use in gene therapy. A polynucleotide or
CC polypeptide of the invention is useful for preventing, treating or
CC ameliorating a medical condition, e.g. a proliferative disorder. They are
CC also useful for treating e.g. liver disease, renal failure, immunological
CC disorders including arthritis and psoriasis, cardiovascular disorders
CC such as congenital heart defects and congestive heart failure, and
CC cancer. A method of the invention is useful for diagnosing a pathological
CC condition or susceptibility to a condition in a subject. The present
CC sequence is used in the exemplification of the invention
XX
SQ Sequence 625 AA;
XX
Alignment Scores:
Pred. No.: 3.53e-10 Length: 625
Score: 20.00 Matches: 20
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.00% Indels: 0
DB: 5 Gaps: 0
US-10-029-345A-108_COPY_532_2532 (1-2001) x ABR52382 (1-625)
OY 730 GTGCAGTGTGATGCGGATCTCCGCTCCGACCATGCGCTATCGCTACATCATGANG 789
Db 244 ValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIleMetLys 263
RESULT 22
ABR52350
ID ABR52350 standard; protein; 625 AA.
XX
AC ABR52350;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein relating to the invention SEQ ID NO: 39.
XX
KW anti-proliferative; hepatotropic; nephrotropic; antiarthritic;
KW anti-psoriatic; cardiant; cytosatic; gene therapy; liver disease;
KW proliferative disorder; renal failure; cardiovascular disorder;
KW immunological disorder; arthritis; psoriasis; congenital heart defect;
KW congestive heart failure; cancer; human; mouse; yeast; rat; fruitfly.
OS Homo sapiens.
XX
PN WO200257460-A2.
XX
PD 25-JUL-2002.
XX
PF

XX 20-DEC-2001; 2001WO-US050459.
XX
PR 20-DEC-2000; 2000US-0256868P.
PR 30-MAR-2001; 2001US-0280186P.
PR 01-MAY-2001; 2001US-0287735P.
PR 05-JUN-2001; 2001US-0295848P.
PR 25-JUN-2001; 2001US-0300465P.
XX
XX (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Jackson DG, Feder J, Nelson T, Mintier G, Ramanathan C, Lee L;
PI Siemers N, Bol D, Schieven G, Finger J, Todderud CG, Basolino D;
PI Kryetek S, Mcatee P, Suchard S, Barnes D;
XX
DR WPI; 2002-599721/64.
XX
XX Novel polynucleotides encoding human phosphatase polypeptides useful in
PT the prevention or treatment of e.g. proliferative and cardiovascular
PT disorders.
XX
XX Disclosure; Fig 10; 801pp; English.
XX
XX The invention relates to a novel isolated nucleic acid comprising a
CC polynucleotide having a nucleotide sequence selected from 40
CC polynucleotides fully defined in the specification. The polynucleotide of
CC the invention has antiproliferative, hepatotropic, nephrotropic, CC
CC antidiabetic, antipruritic, cardiac, and cytostatic activity. The
CC polynucleotide may have a use in gene therapy. A polynucleotide or
CC polypeptide of the invention is useful for preventing, treating or
CC ameliorating a medical condition, e.g. a proliferative disorder. They are
CC also useful for treating e.g. liver disease, renal failure, immunological
CC disorders including arthritis and psoriasis, cardiovascular disorders
CC such as congenital heart defects and congestive heart failure, and
CC cancer. A method of the invention is useful for diagnosing a pathological
CC condition or susceptibility to a condition in a subject. The present
CC sequence is used in the exemplification of the invention
XX
XX Sequence 625 AA;
SO
Alignment Scores:
Pred. No.: 3.53e-10 Length: 625
Score: 20.00 Matches: 20
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.00% Indels: 0
DB: Gaps: 0
US-10-029-345A-108_COPY_532_2532 (1-2001) x ABR52350 (1-625)
QY 730 GTGCACTGTTTACCTGGGATCTCCGCTCCGCCACCATGCGCTATCGCTACATGAG 789
DB 244 ValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIleMetLys 263
RESULT 23
ABG73440
ID ABG73440 standard; protein; 625 AA.
XX
XX ABG73440;
AC
XX
XX 10-MAY-2003 (first entry)
DT
XX
XX Human dual specific phosphatase 8 polypeptide.
DE
XX
XX Human; dual specific phosphatase 8; enzyme; infection; inflammation;
KW tumour formation; cytostatic; antiinflammatory.
XX
XX Homo sapiens.
OS
XX US6482644-B1.
FN
XX 19-NOV-2002.
PD
XX

PF 01-AUG-2001; 2001US-00920668.
XX
XX 01-AUG-2001; 2001US-00920668.
PR
XX (ISTS-) ISTS PHARM INC.
XX
XX Cowser LM;
XX
XX WPI; 2003-298140/29.
DR N-PSDB; ABX10760.
XX
XX
XX New antitense compound targeted to a nucleic acid encoding human dual
PT specific phosphatase 8, for modulating gene expression and treating
PT diseases associated with expression of the phosphatase in humans.
XX
XX Disclosure; Col 49-54; 36pp; English.
XX
XX The invention relates to a compound targeted to the coding region of a
CC nucleic acid encoding human dual specific phosphatase 8, where the
CC compound specifically hybridizes with the region and inhibits the
CC expression of human dual specific phosphatase 8. The compound is useful
CC for inhibiting the expression of human dual specific phosphatase 8 in
CC cells or tissues, and for treating an animal, particularly a human,
CC suspected of having or being prone to a disease or condition associated
CC with expression of dual specific phosphatase 8. The compound is useful
CC for diagnostics, therapeutics and as a research reagent, e.g. to prevent
CC or delay infection, inflammation or tumour formation, and to distinguish
CC between functions of various members of a biological pathway. This
CC sequence represents human dual specific phosphatase 8
XX
XX Sequence 625 AA;
SO
Alignment Scores:
Pred. No.: 3.53e-10 Length: 625
Score: 20.00 Matches: 20
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.00% Indels: 0
DB: Gaps: 0
US-10-029-345A-108_COPY_532_2532 (1-2001) x ABG73440 (1-625)
QY 730 GTGCACTGTTTACCTGGGATCTCCGCTCCGCCACCATGCGCTATCGCTACATGAG 789
DB 244 ValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIleMetLys 263
RESULT 24
AAW29150
ID AAW29150 standard; protein; 663 AA.
XX
XX AAW29150;
AC
XX
XX 15-DEC-1997 (first entry)
DT
XX
XX Dual-specific murine threonine-tyrosine phosphatase M3/6.
DE
XX
XX murine; phosphatase; M3/6; dual specificity; Threonine; Tyrosine;
KW inactivate; mitogen activated protein kinase; MAP-K; cdc25 pty; yeast;
KW trinucleotide repeat; repeat expansion; neurodegenerative; proliferative;
KW diagnosis; tumour; lung; brain; chromosomal deletion.
XX
XX Mus sp.
OS
XX
XX Key Location/Qualifiers
FH Misc-difference 29..49
FT FT /note="region of homology with yeast cdc25"
FT Misc-difference 117..136
FT FT /note="region of homology with yeast cdc25"
FT Domain 244..253
FT /label=catalytic_domain
XX
XX MO9706245-A1.
XX

PD 20-FEB-1997.
 XX
 PF 05-AUG-1996; 96MO-GB001906.
 XX
 PR 04-AUG-1995; 95GB-00016059.
 XX
 PA (MEDI-) MEDICAL RES COUNCIL.
 XX
 PI Davies KE, Theodosiou A;
 XX
 DR WPI; 1997-154253/14.
 N-PSDB; AAT86758.
 XX
 PT Murine mitogen activated protein (MAP) kinase phosphatase M3/6 - is a
 PT suspected dual specificity Thr-Tyr phosphatase; useful for diagnosing and
 PT treating neuro-degenerative or proliferative diseases e.g. tumours.
 XX
 PS Claim 23; Fig 2; 51pp; English.
 XX
 CC This is a murine phosphatase designated M3/6 (encoded by AAT86758). M3/6
 CC is a suspected dual specificity Threonine-Tyrosine phosphatase, capable
 CC of inactivating mitogen activated protein (MAP) kinase. The M3/6 protein
 CC product shows high homology to the cdc25 p1p of yeast at residues 29-49
 CC and 117-136. The gene also contains a complex triplet distal to the
 CC catalytic domain which is translated into the protein. This domain
 CC comprises a run of 4 serine residues which in turn is followed by a
 CC -terminal section by a single asparagine. This makes the phosphatase gene
 CC a candidate for a human disease caused by repeat expansion or mutation.
 CC M3/6 is expressed highly in the brain and may have utility in
 CC investigating signal transduction mechanisms in brain and muscle. The
 CC M3/6 and Hb5 (a human homologue) genes may be responsible, if mutated,
 CC for various neurodegenerative or proliferative diseases, and may
 CC therefore be used for the diagnosis of such diseases, e.g. tumours,
 CC especially lung or brain tumours, associated with deletion of the
 CC chromosomal region 11p15.5. The polypeptides can be used to screen for
 CC inhibitors to treat these diseases
 XX
 SQ Sequence 663 AA;
 XX
 Alignment Scores:
 Pred. No.: 3.51e-10 Length: 663
 Score: 20.00 Matches: 20
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.00% Indels: 0
 DB: Gaps: 0
 XX
 US-10-029-345A-108_COPY_532_2532 (1-2001) x AAM29150 (1-663)
 OY 730 GTGCACCTGTTTACCTGGGATCTCCGCTCCGCCACCATGCTATCGCTACATCATGAAG 789
 Db 244 ValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaThrIleMetLys 263
 RESULT 25
 ID ABR52351
 XX ABR52351 standard; protein; 663 AA.
 AC ABR52351;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Protein relating to the invention SEQ ID NO: 40.
 XX
 KW antiproliferative; hepatotropic; nephrotropic; antiarthritic;
 KW antipsoriatic; cardiant; cyostatic; gene therapy; liver disease;
 KW proliferative disorder; renal failure; cardiovascular disorder;
 KW immunological disorder; arthritis; psoriasis; congenital heart defect;
 KW congestive heart failure; cancer; human; mouse; yeast; rat; fruitfly.
 XX
 OS Mus musculus.
 XX
 PN WO200257460-A2.

XX
 PD 25-JUL-2002.
 XX
 PF 20-DEC-2001; 2001MO-US050459.
 XX
 PR 20-DEC-2000; 2000US-0256686P.
 XX
 PR 30-MAR-2001; 2001US-0280186P.
 PR 01-MAY-2001; 2001US-0287735P.
 PR 05-JUN-2001; 2001US-0295848P.
 PR 25-JUN-2001; 2001US-0300465P.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Jackson DG, Feder J, Nelson T, Mintier G, Ramanathan C, Lee J;
 PI Siemers N, Bol D, Schieven G, Finger J, Todderud CG, Bassolino D;
 PI Kyteck S, Mcatee P, Suchard S, Banas D;
 XX
 DR WPI; 2002-599721/64.
 XX
 PT Novel polynucleotides encoding human phosphatase polypeptides useful in
 PT the prevention or treatment of e.g. proliferative and cardiovascular
 PT disorders.
 XX
 PS Disclosure; Fig 10; 801pp; English.
 XX
 CC The invention relates to a novel isolated nucleic acid comprising a
 CC polynucleotide having a nucleotide sequence selected from 40
 CC polynucleotides fully defined in the specification. The polynucleotide of
 CC the invention has antiproliferative, hepatotropic, nephrotropic, of
 CC antarthritic, antipsoriatic, cardiant, and cyostatic activity. The
 CC polynucleotide may have a use in gene therapy. A polynucleotide or
 CC polypeptide of the invention is useful for preventing, treating or
 CC ameliorating a medical condition, e.g. a proliferative disorder. They are
 CC also useful for treating e.g. liver disease, renal failure, immunological
 CC disorders including arthritis and psoriasis, cardiovascular disorders
 CC such as congenital heart defects and congestive heart failure, and
 CC cancer. A method of the invention is useful for diagnosing a pathological
 CC condition or susceptibility to a condition in a subject. The present
 CC sequence is used in the exemplification of the invention
 XX
 SQ Sequence 663 AA;
 XX
 Alignment Scores:
 Pred. No.: 3.51e-10 Length: 663
 Score: 20.00 Matches: 20
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.00% Indels: 0
 DB: Gaps: 0
 XX
 US-10-029-345A-108_COPY_532_2532 (1-2001) x ABR52351 (1-663)
 OY 730 GTGCACCTGTTTACCTGGGATCTCCGCTCCGCCACCATGCTATCGCTACATCATGAAG 789
 Db 244 ValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaThrIleMetLys 263
 RESULT 26
 ID ABB63527
 XX ABB63527 standard; protein; 476 AA.
 AC ABB63527;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 17373.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.

PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
XX
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR MPI; 2001-656860/75.
XX
DR N-PSDB; ABL07630.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
XX
PT genes from *Drosophila* and for elucidating cell signaling and cell-cell
XX
PT interactions.
XX
PS Disclosure; SEQ ID NO 17373; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
XX
CC capable of detecting 1000 or more genes from *Drosophila*. The invention is
XX
CC useful in developmental biology and in elucidating cell signalling and
XX
CC cell-cell interactions in higher eukaryotes for the development of
XX
CC insecticides, therapeutics and pharmaceutical drugs. The invention
XX
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
XX
CC sequences (AB101840-AB16175) and the encoded proteins (AB57737-
XX
CC AB572072). The sequence data for this patent did not form part of the
XX
CC printed specification, but was obtained in electronic format directly
XX
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 476 AA;
XX
Alignment Scores:
Pred. No.: 0.00281 Length: 476
Score: 13.00 Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.95% Indels: 0
DB: 4 Gaps: 0
XX
US-10-029-345A-108_COPY_532_2532 (1-2001) x ABB63527 (1-476)
QY 742 GCTGGATCTCCCGCTCCGCCACATCGCATGCCCTAC 780
ABG00724
DB 218 AAGIYLleSeraTserAlaTrIleAlaIleAlaIleAtr 230
XX
RESULT 27
ID ABG00724 standard; protein; 836 AA.
XX
AC ABG00724;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #715.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
XX
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX

PI Drmanac RT, Liu C, Tang YF;
XX
XX MPI; 2001-639362/73.
XX
DR N-PSDB; AAG64911.
XX
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
XX
PT diagnostics, forensics, gene mapping, identification of mutations
XX
PT responsible for genetic disorders or other traits and to assess
XX
PT biodiversity.
XX
PS Claim 20; SEQ ID NO 31083; 103pp; English.
XX
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX
CC sequences. (I) is useful as hybridisation probes, polymerase chain
XX
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX
CC and in recombinant production of (II). The polynucleotides are also used
XX
CC in diagnostics as expressed sequence tags for identifying expressed
XX
CC genes. (I) is useful in gene therapy techniques to restore normal
XX
CC activity of (II) or to treat disease states involving (II). (II) is
XX
CC useful for generating antibodies against it, detecting or quantitating a
XX
CC polypeptide in tissue, as molecular weight markers and as a food
XX
CC supplement. (II) and its binding partners are useful in medical imaging
XX
CC of sites expressing (II). (I) and (II) are useful for treating disorders
XX
CC involving aberrant protein expression or biological activity. The
XX
CC polypeptide and polynucleotide sequences have applications in
XX
CC diagnostics, forensics, gene mapping, identification of mutations
XX
CC responsible for genetic disorders or other traits to assess biodiversity
XX
CC and to produce other types of data and products dependent on DNA and
XX
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
XX
CC amino acid sequences of the invention. Note: The sequence data for this
XX
CC patent did not appear in the printed specification, but was obtained in
XX
CC electronic format directly from WIPO at
XX
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 836 AA;
XX
Alignment Scores:
Pred. No.: 0.0251 Length: 836
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: 4 Gaps: 0
XX
US-10-029-345A-108_COPY_532_2532 (1-2001) x ABG00724 (1-836)
QY 724 GTTCTAGTGCACCTGTTAGCTGGATCTCCGCTCC 759
ABU29344
DB 744 ValLeuValHleCysLeuHlaGlyIleSeraTser 755
XX
RESULT 28
ID ABU29344 standard; protein; 498 AA.
XX
AC ABU29344;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by prokaryotic essential gene #14871.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
KW Enterococcus faecalis.
XX
OS
XX
PN WO200271183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
XX
PR 06-SEP-2001; 2001US-00948993.
XX
PR 25-OCT-2001; 2001US-0342923P.
XX

PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Hasebeck R, Ohlsen KL, Zykkind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
DR N-PSDB; ACA33214.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
XX Claim 25; SEQ ID NO 57268; 1766bp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 498 AA;
XX
XX Alignment Scores:
XX Pred. No.: 2 49
XX Score: 10.00 Length: 498
XX Percent Similarity: 100.00% Matches: 10
XX Best Local Similarity: 100.00% Conservative: 0
XX Query Match: 1.55% Mismatches: 0
XX DB: Indels: 0
XX Gaps: 0
XX
US-10-029-345A-108_COPY_532_2532 (1-2001) x ABU29344 (1-498)
OY 103 GCGGCTATCATTTAGCAGACTTTTCCG 74
DB 342 AAGAGTGTGlnGlnLeuAlaLeuPhePro 351
XX
RESULT 29
IDA44807
XX ADA44807 standard; protein; 473 AA.
AC ADA44807;
XX
XX 04-DEC-2003 (first entry)
XX

DE CD4/TCR CD3epsilon chain chimeric protein CD4epsilon1015, SEQ ID NO:2.
XX
XX HIV-1 infection; human immunodeficiency virus-1; CD4+ cell; chimeric CD4;
XX endoplasmic reticulum; ER retention; envelope protein gp160;
XX T cell receptor CD3epsilon chain; C-terminal domain; CD4epsilon1015;
XX gene therapy; human; receptor.
XX
XX OS Chimeric.
XX OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX FT 1..458
XX FT /label= CD4
XX FT 459..473
XX FT Region
XX FT /note= "Part of the C-terminal domain of the T cell
XX receptor CD3epsilon chain"
XX
XX WO2003076468-A1.
XX
XX PD 18-SEP-2003.
XX
XX PD 14-MAR-2003; 2003WO-ES000120.
XX
XX PF 14-MAR-2002; 2002ES-00000616.
XX
XX PR (CNSG) CONSENSO SUPERIOR INVESTIGACIONES CIENTIF.
XX
XX PA Alarcon Sanchez BJ, San Jose Martinez ME, Zaldivar Notario I;
XX PI Gomez Buendia M;
XX DR WPI; 2003-779059/73.
XX N-PSDB; ADA44806.
XX
XX PT Composition for treating or preventing human immune deficiency virus,
XX PT comprises CD4 chimeric protein having a protective effect in trans, or
XX PT related nucleic acid.
XX
XX PS Claim 5; Page 33-35; 43pp; Spanish.
XX
XX CC The invention relates to a composition for the treatment or prevention of
XX CC human immunodeficiency virus-1 (HIV-1) infection. The composition
XX CC comprises CD4+ cells that have been transduced with a vector that encodes
XX CC a chimeric CD4 molecule which is capable of being retained in the
XX CC endoplasmic reticulum (ER). The invention also encompasses the use of a
XX CC soluble protein factor produced by CD4+ cells that have been transduced
XX CC with a vector encoding a chimeric CD4 protein; and the use of an
XX CC expression system encoding a chimeric CD4 protein; and the use of an
XX CC chimeric CD4 molecule binds to the HIV-1 envelope protein gp160,
XX CC resulting in HIV-1 retention in the ER and thereby preventing viral
XX CC replication. In a specific embodiment, the chimeric CD4 molecule
XX CC comprises CD4 fused to 15 amino acids of the C-terminal domain of the T
XX CC cell receptor CD3epsilon chain; this chimeric CD4 molecule is designated
XX CC CD4epsilon1015 (ADA44807). A known chimeric CD4 of similar structure but
XX CC containing only 10 amino acids from CD3epsilon chain can also be used.
XX CC Compositions of the invention have an in trans effect on the replication
XX CC of HIV-1, and may be used to treat and prevent HIV-1 infection. The
XX CC present sequence represents the chimeric CD4 molecule CD4epsilon1015, which
XX CC is specifically claimed for use in compositions of the invention.
XX
XX SQ Sequence 473 AA;
XX
XX Alignment Scores:
XX Pred. No.: 232
XX Score: 8.00 Length: 473
XX Percent Similarity: 100.00% Matches: 8
XX Best Local Similarity: 100.00% Conservative: 0
XX Query Match: 1.20% Mismatches: 0
XX DB: Indels: 0
XX Gaps: 0
XX
US-10-029-345A-108_COPY_532_2532 (1-2001) x ADA44807 (1-473)
OY 1819 GACAGAGTGTGACTGCGCGGAGC 1842
XX

Db 78 AspArgAlaAspSerArgArgSer 85

RESULT 30

AAV59170
ID AAV59170 standard; protein; 474 AA.
XX
AC AAV59170;
XX
DT 14-MAR-2000 (first entry)
XX
DE CD4-Ig fusion protein CD4Mmu.
XX
DE HIV; extracellular; CD4; gp120; immunoglobulin; Ig; fusion protein;
KM secreted protein; SIV infection; medicament.
XX
OS Synthetic.
XX Homo sapiens.
XX
PN CAL340741-C.
XX
PD 14-SEP-1999.
XX
PF 20-JAN-1989; 89CA-00588749.
XX
PR 20-JAN-1989; 89CA-00588749.
XX
PA (GEHO) GEN HOSPITAL CORP.
XX
PI Seed B;
XX
DR WPI; 2000-063015/06.
XX
DR N-PSDB; AAZ48203.
XX
PT New fusion gene encoding immunoglobulin-CD4 fusion proteins, useful in
PS the treatment of HIV or simian immunodeficiency virus infections.
XX
PS Example 1; Page 47-53; 89pp; English.
XX
CC The invention provides a fusion gene encoding a fusion protein that
CC comprises an extracellular CD4 DNA sequence or its fragment which binds
CC to HIV gp120 when fused to an immunoglobulin (Ig) chain and the DNA
CC sequence of an Ig heavy or light chain, where the DNA sequence encoding
CC the variable region has been replaced with the DNA sequence which encodes
CC extracellular CD4 or its gp120 binding fragment. The fusion protein is
CC capable of being secreted. The fusion proteins are useful for treating
CC HIV or SIV infections in animals, preferably humans. They are also useful
CC for producing medicaments which can be used for treating HIV or SIV
CC infections in humans. The present sequence represents the fusion protein
CC CD4Mmu where the CD4 is linked to human IgG1 at the Met2 site upstream of
CC the CH1 region
XX
SQ Sequence 474 AA;

Alignment Scores:

Pred. No.:	232	Length:	474
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.20%	Indels:	0
DB:	3	Gaps:	0

US-10-029-345A-108_COPY_532_2532 (1-2001) x AAV59170 (1-474)

QY 1819 GACAGAGCTGACTCGCGCGGAGC 1842
|||||
DB 78 AspArgAlaAspSerArgArgSer 85

RESULT 31

AAV93011
ID AAV93011 standard; protein; 481 AA.
XX
AC AAV93011;
XX

DT 25-MAR-2003 (revised)
DT 03-AUG-1992 (first entry)
XX
XX Genetic construct which encodes CD4 linked to human IGM at the Met site
DE upstream of the CH2 region (fusion protein CD4Pmu).
XX
XX Fusion protein; immunoglobulin-like molecule; HIV; SIV; therapy;
KM diagnosis; CD4; gp120; binding fragment; glycoprotein; variable region.
XX
OS Homo sapiens.
XX
PN EP325262-A.
XX
PD 26-JUL-1989.
XX
PF 20-JAN-1989; 89EP-00100913.
XX
PR 22-JAN-1988; 88US-00147351.
XX
PA (GEHO) GEN HOSPITAL CORP.
XX
PI Seed B;
XX
DR WPI; 1989-214472/30.
XX
DR N-PSDB; AAN90359.
XX
PT Immunoglobulin-CD4 fusion proteins - used for treating HIV or SIV
PS infections or detecting HIV or SIV in sample.
XX
PS Example; Table 4, Page 41-47; 68pp; English.
XX
CC The fusion protein genes of the invention pref. comprises cDNA sequences
CC which encode CD4 or a fragment which binds gp120 ligated to an expression
CC plasmid which encodes an antibody in which the variable region of the
CC gene has been deleted (see WO87-02671). The CD4 portion of the fusion
CC protein may comprise the complete CD4 sequence, the 370 AA extracellular
CC region and the membrane spanning domain, or the extracellular region. The
CC Ig heavy chain is pref. from IGM, IgG1 or IgG3. The following are
CC specifically claimed: fusion proteins CD4H1ambda1, CD4Mmu, CD4Pmu,
CC CD4E1ambda1, and CD4Mmu (No. 67608), pCD4E1ambda (No. 67609) and
CC pCD41ambda1 (No. 67610). (Updated on 25-MAR-2003 to correct PA field.)
XX
SQ Sequence 481 AA;

Alignment Scores:

Pred. No.:	232	Length:	481
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.20%	Indels:	0
DB:	1	Gaps:	0

US-10-029-345A-108_COPY_532_2532 (1-2001) x AAV93011 (1-481)

QY 1819 GACAGAGCTGACTCGCGCGGAGC 1842
|||||
DB 78 AspArgAlaAspSerArgArgSer 85

RESULT 32

AAV19510
ID AAV19510 standard; protein; 481 AA.
XX
AC AAV19510;
XX
DT 09-JAN-2001 (first entry)
XX
DE CD4-IgM fusion protein CD4Pmu.
XX
XX CD4; IGM, human; CD4Pmu; fusion protein; immunoglobulin; HIV; SIV; gp120;
KM therapy; diagnosis.
XX
OS Homo sapiens.
XX


```

PR 20-JAN-1989; 89CA-00588749.
XX
PA (GEHO ) GEN HOSPITAL CORP.
XX
PI Seed B;
XX
DR WPI; 2000-063015/06.
XX
DR N-PSDB; AAZ48204.
XX
PT New fusion gene encoding immunoglobulin-CD4 fusion proteins, useful in
XX the treatment of HIV or simian immunodeficiency virus infections.
XX
PS Example 1; Page 54-60; 89pp; English.
XX
CC The invention provides a fusion gene encoding a fusion protein that
XX comprises an extracellular CD4 DNA sequence or its fragment which binds
XX to HIV gp120 when fused to an immunoglobulin (Ig) chain and the DNA
XX sequence of an Ig heavy or light chain, where the DNA sequence encoding
XX the variable region has been replaced with the DNA sequence which encodes
XX extracellular CD4 or its gp120 binding fragment. The fusion protein is
XX capable of being secreted. The fusion proteins are useful for treating
XX HIV or SIV infections in animals, preferably humans. They are also useful
XX for producing medicaments which can be used for treating HIV or SIV
XX infections in humans. The present sequence represents the fusion protein
XX cDNA where the CD4 is linked to human IgG1 at the Pst site upstream of
XX the CH2 region
XX
SQ Sequence 481 AA;
XX
Alignment Scores:
Pred. No.: 232
Score: 8.00 Length: 481
Percent Similarity: 100.00% Matches: 8
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
Gaps: 0
US-10-029-345A-108_COPY_532_2532 (1-2001) x AAY59171 (1-481)
OY 1819 GACAGAGCTGACTGCGCGCGGAGC 1842
DB 78 AsparthalapSerargArgSer 85
|||||
RESULT 35
ID ADE08675
AC ADE08675 standard; protein; 502 AA.
XX
XX ADE08675;
XX
XX 29-JAN-2004 (first entry)
XX
XX
XX Novel protein (useful for identifying genetic disorders) #830.
XX
XX novel gene; novel protein; tissue marker; molecular weight marker;
XX chromosome marker; genetic disorder.
XX
XX Unidentified.
XX
XX OS
XX WO2003054152-A2.
XX
XX
XX 03-JUL-2003.
XX
XX 10-DEC-2002; 2002WO-US039555.
XX
XX 10-DEC-2001; 2001US-0339739P.
XX 11-DEC-2001; 2001US-0339453P.
XX 14-MAR-2002; 2002US-0365081P.
XX 14-MAR-2002; 2002US-0365384P.
XX 12-APR-2002; 2002US-0372381P.
XX 12-APR-2002; 2002US-0372615P.
XX 22-APR-2002; 2002US-00128558.
XX 24-APR-2002; 2002US-0376045P.
XX

```

PA (HYSE-)HYSEQ INC.
 XX
 PI Tang YT, Asundi V, Goodrich RM, Ren F, Zhang J, Zhao QA, Wang J;
 PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Dymnac RT, Wang Z;
 PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;
 XX
 DR WPI: 2003-569235/53.
 DR N-PSDB: ADE07764.
 XX
 XX New polynucleotides, useful for expressing recombinant proteins for
 PT analysis, characterization or therapeutic use, or as markers for tissues
 PT in which the corresponding protein is preferentially expressed.
 XX
 PS Claim 20; SEQ ID NO 1741; 1177bp; English.
 XX
 CC The invention comprises the amino acid and coding sequences of novel
 CC proteins. The DNA and protein sequences of the invention are useful as:
 CC markers for tissues in which the corresponding protein is preferentially
 CC expressed; as molecular weight markers on gels; as chromosome markers on
 CC tags; to identify chromosomes or to map related gene positions; and to
 CC compare with endogenous DNA sequences in patients to identify potential
 CC genetic disorders. The present amino acid sequence represents a protein
 CC of the invention.
 XX
 SQ Sequence 502 AA;
 XX
 Alignment Scores:
 Pred. No.: 231 Length: 502
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.24% Indels: 0
 DB: 7 Gaps: 0
 XX
 US-10-029-345A-108_COPY_532_2532 (1-2001) x ADE08675 (1-502)
 OY 1493 TGGGCGGTGCCACTGCTGCTGCTT 1470
 DB 365 TrpAlaValProLeuLeuVal 372
 XX
 RESULT 36
 ID AAB00158 standard; protein; 507 AA.
 XX
 AC AAB00158;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE sCD4-SCFv(17b) HIV single chain antibody fusion protein.
 XX
 KW Fusion protein; HIV; human immunodeficiency virus; antibody; Fv; AIDS;
 KW acquired immune deficiency syndrome; neutralisation; infection;
 KW gene therapy; CD4; gp120; glycoprotein; resistance; vaccination;
 KW binding domain; single chain antibody; chimera; chimeric protein.
 XX
 OS Human immunodeficiency virus.
 OS Synthetic.
 OS
 PN WO200055207-A1.
 PD 21-SEP-2000.
 PP 16-MAR-2000; 2000WO-US006946.
 PR 16-MAR-1999; 99US-0124681P.
 PA (USSH) US NAT INST OF HEALTH.
 PI Berger EA, Del Castillo CM;
 DR WPI: 2000-638183/61.
 DR N-PSDB: AAA54045.
 XX

PT Novel neutralizing bispecific fusion proteins effective in viral such as
 PT HIV neutralization, comprises two different binding domains, inducing-
 PS binding domain and induced-binding domain functionally linked by linker.
 XX Claim 39; Page 46-47; 55pp; English.

CC sCD4-SCFv(17b) is a neutralising bispecific fusion protein capable of
 CC binding to two sites of its target protein. The protein comprises a first
 CC binding domain capable of binding to an inducing site on the target
 CC protein, a second binding domain capable of forming neutralising complex
 CC with an induced epitope of the target protein and a linker connecting the
 CC binding domains. sCD4-SCFv(17b) comprises a soluble CD4 fragment
 CC (containing domains D1 and D2) fused to a single chain Fv portion of
 CC antibody 17b via a linker. sCD4-SCFv(17b), its variant, analogue or
 CC mimetic is used for inactivating gp120 protein of HIV, and for
 CC neutralising HIV. It is also used for blocking and preventing the binding
 CC of the viral or recombinant gp120 protein to soluble CD4 or lymphocyte
 CC CD4 and for inhibiting HIV replication. The chimeric proteins is
 CC therefore useful for treating HIV infection and also AIDS. It is are
 CC particularly useful in the prevention of infection during or immediately
 CC after HIV exposure (e.g., mother/infant transmission, post-exposure
 CC prophylaxis, and as a topical inhibitor) and for providing long term
 CC resistance to HIV infections and AIDS. Gene therapy is used to secrete
 CC the bispecific protein at mucosal surfaces, such as the vaginal, rectal
 CC or oral mucosa. The fusion proteins is highly potent, broadly cross-
 CC reactive with neutralising antibody with high in vivo activity and no Fc-
 CC mediated undesirable labelling properties. When the fusion protein is
 CC substantially derived from human proteins, it has minimal immunogenicity
 CC and toxicity in humans which is of great value in prevention of infection
 CC during or immediately after HIV exposure
 XX

SQ Sequence 507 AA;

Alignment Scores:
 Pred. No.: 230 Length: 507
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.20% Indels: 0
 DB: 3 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x AAB00158 (1-507)

OY 1819 GACAGAGCTGACTCGCGCGGAGC 1842
 DB |||||
 78 AsparagalaespaSerArgSer 85

RESULT 37

AAW26792 ID AAW26792 standard; protein; 509 AA.

AC AAW26792;

DT 22-JUN-1998 (first entry)

DE Mouse TIE-2 receptor ligand-3.

XX TIE-2 receptor ligand; TIE ligand-3; tyrosine kinase; mouse;

KW angiogenesis; vascularisation; blood vessel growth; tumour; therapy;

KW receptorbody.

OS Mus musculus.

PN WO9748804-A2.

XX 24-DEC-1997.

PD 19-JUN-1997; 97WO-US010728.

PF 19-JUN-1996; 96US-00665926.

PR 02-JUL-1996; 96US-0021087P.

PR 02-AUG-1996; 96US-0022999P.

XX

PA (REG-) REGENERON PHARM INC.

XX Valenzuela DM, Jones PF, Yancopoulos GD;

XX WPI; 1998-063143/06.

DR N-PSDB; AAT99594.

PT DNA encoding TIE ligands 3 and 4 - useful for, e.g. blocking blood vessel

XX growth and promoting neovascularisation, etc.

PS Example 10; Fig 6A-B; 86pp; English.

XX This protein comprises novel mouse TIE ligand-3 (mTL-3), a TIE-2 receptor
 CC ligand that is a tyrosine kinase with immunoglobulin and epidermal growth
 CC factor homology domains. Its amino acid sequence was deduced from an
 CC isolated cDNA clone (see AAT99594). A host-vector system for the
 CC production of mTL-3 is claimed, as well as an antibody which specifically
 CC binds mTL-3, a receptorbody which specifically binds mTL-3, a conjugate
 CC comprising mTL-3 and a cytotoxic agent, and an isolated nucleic acid (see
 CC AAT99593) coding for human TIE ligand 4 (see AAW26791). TL3, TL4, their
 CC antagonists, etc. are useful for blocking blood vessel growth, promoting
 CC neovascularisation, promoting or blocking the growth or differentiation
 CC of a cell expressing the TIE receptor and attenuating or preventing
 CC tumour growth in humans
 XX

SQ Sequence 509 AA;

Alignment Scores:
 Pred. No.: 230 Length: 509
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.20% Indels: 0
 DB: 2 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x AAW26792 (1-509)

OY 1083 CGTGCCAGCGTGCCGCGCGTGA 1106
 DB |||||
 86 ArgAlaGlnArgAlaGlnArgAla 93

RESULT 38

AAW47533 ID AAW47533 standard; protein; 509 AA.

AC AAW47533;

DT 09-SEP-1998 (first entry)

DE Amino acid sequence of murine TIE ligand 3.

XX Chimeric TIE ligand; TIE-2 ligand; neovascularisation; tumour; mouse.

XX Mus sp.

PN WO9805779-A1.

XX 12-FEB-1998.

PD 01-AUG-1997; 97WO-US013557.

PF 02-AUG-1996; 96US-0022999P.

PR 25-OCT-1996; 96US-00740223.

XX (REG-) REGENERON PHARM INC.

PA Davis S, Yancopoulos GD;

XX WPI; 1998-145615/13.

DR N-PSDB; AAV18620.

PT Modified human TIE-2 receptor ligand(s) - useful for promoting wound

PT healing.

XX Example 20; Fig 21; 202pp; English.
PS
XX This is the amino acid sequence of the murine TIE ligand 3, used in the
CC method of the invention, involving the production of TIE-2 ligands which
CC promote healing. The nucleic acids, vectors and host cells used in the
CC method of the invention are useful for the recombinant production of the
CC ligands. The ligands, etc. are useful for blocking blood vessel growth,
CC promoting neovascularisation, promoting the growth or differentiation of
CC a cell expressing the TIE receptor, blocking the growth or differentiation of
CC differentiation of a cell expressing the TIE receptor and for attenuating
CC or preventing tumour growth in a human
XX
SQ Sequence 509 AA;
Alignment Scores:
Pred. No.: 230 Length: 509
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 2 Gaps: 0
US-10-029-345A-108_COPY_532_2532 (1-2001) x AAM47533 (1-509)
QY 1083 CGTGGCCAGCGTGCACCGGTGCA 1106
DB 86 ArgAlaGlnArgAlaGlnArgAla 93
RESULT 39
ID AAY23734 standard; protein; 509 AA.
XX AAY23734;
XX 08-SEP-1999 (first entry)
XX
DE TIE ligand-3 (TL3) amino acid sequence.
XX
KM TIE ligand-3; TL3; angiogenic factor; receptor tyrosine kinase; AR-1;
KM angiogenesis regulator; neoplastic disease; tumour angiogenesis;
KM wound healing; thromboembolic disease; atherosclerosis;
KM inflammatory disease; AR-1 receptor.
XX
OS Mus sp.
XX
PN WO9932639-A1.
XX
PD 01-JUL-1999.
XX
PF 17-DEC-1998; 98WO-US026800.
XX
PR 19-DEC-1997; 97US-00994309.
XX
PA (REG-) REGENERON PHARM INC.
XX
PI Valenzuela DM;
XX
XX WPI; 1999-418933/35.
DR N-PSDB; AAX85781.
XX
XX Receptor tyrosine kinase, AR-1 and related nucleic acid molecules.
PT
XX
PS Example 1; Fig 1A-C; 50pp; English.
XX
XX The present sequence represents TIE ligand-3 (TL3), an angiogenic factor.
CC The specification describes a receptor tyrosine kinase, designated AR-1,
CC which is related to TIE ligands. AR-1 is believed to be a regulator of
CC angiogenesis and thus the factor, as well as nucleic acids encoding it,
CC are useful in the diagnosis and treatment of certain diseases such as
CC neoplastic diseases involving tumour angiogenesis, wound healing,
CC thromboembolic diseases, atherosclerosis and inflammatory diseases. AR-1
CC can also be used to support the survival and/or growth and/or migration

CC and/or differentiation of human AR-1 receptor expressing cells. AR-1 can
CC be used to identify the AR-1 receptor. AR-1-cytotoxic conjugates can be
CC used to target tumours expressing AR-1 or its receptor. Antagonists of AR
CC -1, e.g. antibodies are useful to prevent or attenuate its biological
CC activity
XX
SQ Sequence 509 AA;
Alignment Scores:
Pred. No.: 230 Length: 509
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 2 Gaps: 0
US-10-029-345A-108_COPY_532_2532 (1-2001) x AAY23734 (1-509)
QY 1083 CGTGGCCAGCGTGCACCGGTGCA 1106
DB 86 ArgAlaGlnArgAlaGlnArgAla 93
RESULT 40
ID AAY90398 standard; protein; 509 AA.
XX AAY90398;
XX 12-FEB-2001 (first entry)
XX
DE Mouse TIE ligand-3 protein sequence.
XX
KM AR-2 fibrinogen-like domain; angiotensin related-2; mouse; ischemia;
KM diabetes; tumour angiogenesis; neoplastic disease; atherosclerosis;
KM thromboembolic disease; inflammatory disease; wound healing;
KM vascularisation; therapy; diagnosis; TIE ligand-3.
XX
OS Mus sp.
XX
PN WO200052167-A1.
XX
PD 08-SEP-2000.
XX
PF 10-FEB-2000; 2000WO-US003381.
XX
PR 02-MAR-1999; 99US-0122499P.
XX
PA (REG-) REGENERON PHARM INC.
XX
PI Jones PF, Valenzuela DM;
XX
XX WPI; 2000-579286/54.
DR N-PSDB; AAA37838.
XX
XX Novel nucleic acid molecules encoding fibrinogen-like domain of
PT Angiotensin Related-2 factor useful for treating ischemia, diabetes, and
PT for wound healing.
XX
XX Example 1; Fig 1; 55pp; English.
XX
XX This sequence is the mouse TIE ligand-3. This sequence was used in the
CC isolation of the angiotensin related-2 (AR-2) fibrinogen-like domain
CC (FD) of the invention. AR-2 is useful for treating and diagnosing
CC ischaemia, diabetes, tumour angiogenesis, neoplastic disease,
CC thromboembolic diseases, atherosclerosis, inflammatory diseases, and for
CC wound healing. The DNA is useful for developing ligands, screening
CC agonists and antagonists of AR-2, and as a therapeutic for treating
CC disorders involving cells, tissues or organs expressing AR-2 receptor. AR
CC -2 is useful to promote the growth, survival, migration, stabilisation or
CC destabilisation, and/or differentiation of cells expressing AR-2
CC receptor. AR-2 is also useful in assay systems to identify agonists and
CC antagonists of AR-2 receptor. AR-2 is also useful for inducing or
CC preventing vascularisation in diseases or disorders where such function

CC is indicated, for delivering toxins to a receptor bearing cells and as
CC diagnostic reagents for detecting the disease by tissue staining or whole
CC body imaging
XX
SQ Sequence 509 AA;

Alignment Scores:

Pred. No.: 230 Length: 509
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 3 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x AAY90398 (1-509)

QY 1083 CGTGGCCAGCGTGCAGCGTGCA 1106
Db 86 ArgAlaGlnArgAlaGlnArgAla 93

RESULT 41

ADD69365
ID ADD69365 standard; protein; 509 AA.

AC ADD69365;

XX 15-JAN-2004 (first entry)

DE Human Ang-3 protein - SEQ ID 101.

XX fusion protein; fibrinogen-like; coiled-like domain;
KW angiotensin-related factor; ARF; angiotensin; Ang-1; Ang-2; Ang-3;
KW Ang-4; Ang-2X; vlnetary; antiinflammatory; vasotropic; necrosis;
KW ischaemia; inflammation; wound healing; CCD; FLD; human.

XX Homo sapiens.

OS WO2003048185-A2.

XX 12-JUN-2003.

XX 21-NOV-2002; 2002WO-US037660.

XX 30-NOV-2001; 2001US-0334488P.

XX (GENV-) GENVEC INC.

XX Kessler PD, Kovacs I;

XX WPI; 2003-513736/48.

XX New fusion protein comprising a fibrinogen-like or coiled-like domain,
PT useful for preparing a composition for treating necrosis, ischaemia or
PT inflammation, or for promoting wound healing. The current

PS Disclosure; SEQ ID NO 101; 340pp; English.

XX The invention relates to a novel fusion protein comprising a fibrinogen-
CC like domain (FLD) or coiled-coil domain (CCD). The domain may be
CC identical or homologous to that of an angiotensin-related factor (ARF),
CC examples of which include Ang (angiotensin)-1, Ang-2, Ang-3, Ang-4 and
CC Ang-2X. The molecules of the invention demonstrate vulnerability,
CC antiinflammatory and vasotropic activities whilst the fusion protein may
CC be useful for preparing a composition for treating necrosis, ischaemia or
CC inflammation, as well as for promoting wound healing. The current
CC sequence is that of the human Ang-3 protein of the invention.

XX Sequence 509 AA;

Alignment Scores:

Pred. No.: 230 Length: 509
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 7 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x ADD69365 (1-509)

QY 1083 CGTGGCCAGCGTGCAGCGTGCA 1106
Db 86 ArgAlaGlnArgAlaGlnArgAla 93

RESULT 42

AAR20152
ID AAR20152 standard; protein; 519 AA.

AC AAR20152;

XX 25-MAR-2003 (revised)

DT 31-MAR-1992 (first entry)

XX Human CD4 sequence encoded by PATY.6.

XX Human immunodeficiency virus; HIV; gp 120; AIDS; ARC; glycoprotein;
KW acquired immune deficiency syndrome; AIDS related complex;
KW T helper lymphocytes.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..25
FT /label= signal_sequence

XX MO9118618-A.

XX 12-DEC-1991.

XX 25-MAY-1990; 90US-00529186.

XX 25-MAY-1990; 90US-00529186.

XX (BIOJ) BIOGEN INC.

XX Fisher RA, Hession C, Burkly LC;

XX WPI; 1992-007200/01.

XX N-PDB; AAQ20327.

XX New immuno-therapeutic human CD4 variants and derivs. - elicit AB
PT production to HIV gp.120, useful in treating, preventing and diagnosing
PT AIDS, ARC and HIV infections.

XX Disclosure; Fig 28; 179pp; English.

XX The sequence was deduced from the DNA sequence of subclone PATY.6, contg.
CC DNA coding for the full-length human CD4. The clone was constructed from
CC plasmids pBG179A and pBG378 (both in US8602940). The DNA can be used to
CC express recombinant CD4 and analogues for use in diagnosis and treatment
CC of diseases caused by infective agents whose primary targets are T4+
CC lymphocytes. See also AAR20148-R20155 and AAR21078. (Updated on 25-MAR-
CC 2003 to correct PA field.)

XX Sequence 519 AA;

Alignment Scores:

Pred. No.: 229 Length: 519
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 2 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x AAR20152 (1-519)

QY 1819 GACAGAGCTGACTCGCGCGGAGC 1842


```
Db      139 AspaRgAlaSpSerArgSer 146
|||||
RESULT 43
AAP94703
ID      AAP94703 standard; protein; 524 AA.
XX
XX      AAP94703;
AC
XX      25-MAR-2003 (revised)
DT      22-MAR-1991 (first entry)
XX
DE      Sequence encoded by T4 lymphocyte cDNA obtained from PBL clone lambda-203
DE      -4.
XX
XX      HIV; soluble T4; immunotherapeutic; prophylactic; diagnostic; AIDS; ARC.
XX
XX      Homo sapiens.
XX
XX      Key      Location/Qualifiers
FH      Misc-difference 67 /note= "AA DESIGNATED NUMBER -23"
FT      Misc-difference 90
FT      Misc-difference 92 /note= "AA DESIGNATED NUMBER 1"
FT      Misc-difference 92 /note= "MATURE N-TERMINUS"
FT
XX      WO8901940-A.
XX
XX      09-MAR-1989.
XX
XX      01-SEP-1988; 88WO-US002940.
XX
XX      04-SEP-1987; 87US-00094322.
XX      07-JAN-1988; 88US-00141649.
XX
XX      (BIOJ ) BIOGEN INC.
XX
XX      Fisher RA, Gilbert W, Sato VL, Flavell RA, Marganore JM;
XX      WPI; 1989-085519/11.
XX      N-PSDB; AAN90642.
XX
XX      DNA sequences coding for soluble T4-like polypeptide(s) - used in
XX      immuno:therapeutic and immunosuppressive compns. and for preventing,
XX      treating or detecting AIDS.
XX
XX      Disclosure; Fig 3; 207pp; English.
XX
XX      The polypeptides encoded are useful in immunotherapeutic, prophylactic
XX      and diagnostic compns. They can be used to purify HIV from a sample. The
XX      soluble T4 protein-based compns. are useful in treating immunodeficient
XX      patients suffering from diseases caused by agents whose primary targets
XX      are T4+ lymphocytes. They can be used for preventing, treating or
XX      detecting AIDS, ARC and HIV infection. (Updated on 25-MAR-2003 to correct
XX      PR field.)
XX
XX      Sequence 524 AA;

Alignment Scores:
Pred. No.:      229      Length:      524
Score:          8.00      Matches:      8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    1.20%      Indels:      0
DB:             1      Gaps:      0

US-10-029-345a-108_COPY_532_2532 (1-2001) x AAP94703 (1-524)

Qy      1819 GACAGAGCTGACTGCGCGCGAGC 1842
      |||||
      144 AspaRgAlaSpSerArgSer 151
```

```
RESULT 44
AAG51289
ID      AAG51289 standard; protein; 525 AA.
XX
XX      AAG51289;
AC
XX      18-OCT-2000 (first entry)
DT
XX
DE      Arabidopsis thaliana protein fragment SEQ ID NO: 65080.
XX
XX      Protein identification; signal transduction pathway; metabolic pathway;
XX      hybridization assay; genetic mapping; gene expression control; promoter;
XX      termination sequence.
XX
XX      Arabidopsis thaliana.
XX
XX      EP1033405-A2.
XX
XX      06-SEP-2000.
XX
XX      25-FEB-2000; 2000EP-00301439.
XX
XX      25-FEB-1999; 99US-0121825P.
XX      05-MAR-1999; 99US-0123180P.
XX      09-MAR-1999; 99US-0123548P.
XX      23-MAR-1999; 99US-0125788P.
XX      25-MAR-1999; 99US-0126785P.
XX      29-MAR-1999; 99US-0126785P.
XX      01-APR-1999; 99US-0127462P.
XX      06-APR-1999; 99US-0128334P.
XX      08-APR-1999; 99US-0128714P.
XX      16-APR-1999; 99US-0129845P.
XX      19-APR-1999; 99US-0130077P.
XX      21-APR-1999; 99US-0130449P.
XX      23-APR-1999; 99US-0130510P.
XX      28-APR-1999; 99US-0130891P.
XX      30-APR-1999; 99US-0132048P.
XX      30-APR-1999; 99US-0132407P.
XX      04-MAY-1999; 99US-0132484P.
XX      05-MAY-1999; 99US-0132485P.
XX      06-MAY-1999; 99US-0132486P.
XX      06-MAY-1999; 99US-0132487P.
XX      07-MAY-1999; 99US-0132863P.
XX      11-MAY-1999; 99US-0134256P.
XX      14-MAY-1999; 99US-0134218P.
XX      14-MAY-1999; 99US-0134219P.
XX      14-MAY-1999; 99US-0134221P.
XX      14-MAY-1999; 99US-0134370P.
XX      18-MAY-1999; 99US-0134768P.
XX      19-MAY-1999; 99US-0134941P.
XX      20-MAY-1999; 99US-0135124P.
XX      21-MAY-1999; 99US-0135353P.
XX      24-MAY-1999; 99US-0135629P.
XX      25-MAY-1999; 99US-0136021P.
XX      27-MAY-1999; 99US-0136392P.
XX      28-MAY-1999; 99US-0136782P.
XX      01-JUN-1999; 99US-0137528P.
XX      03-JUN-1999; 99US-0137528P.
XX      04-JUN-1999; 99US-0137502P.
XX      07-JUN-1999; 99US-0137724P.
XX      08-JUN-1999; 99US-0138094P.
XX      10-JUN-1999; 99US-0138540P.
XX      10-JUN-1999; 99US-0138847P.
XX      14-JUN-1999; 99US-0139119P.
XX      16-JUN-1999; 99US-0139452P.
XX      16-JUN-1999; 99US-0139453P.
XX      17-JUN-1999; 99US-0139492P.
XX      18-JUN-1999; 99US-0139454P.
XX      18-JUN-1999; 99US-0139455P.
XX      18-JUN-1999; 99US-0139456P.
XX      18-JUN-1999; 99US-0139457P.
XX      18-JUN-1999; 99US-0139458P.
```

PR 18-JUN-1999; 99US-0139459P.
PR 18-JUN-1999; 99US-0139460P.
PR 18-JUN-1999; 99US-0139461P.
PR 18-JUN-1999; 99US-0139462P.
PR 18-JUN-1999; 99US-0139463P.
PR 18-JUN-1999; 99US-0139750P.
PR 18-JUN-1999; 99US-0139763P.
PR 21-JUN-1999; 99US-0139817P.
PR 22-JUN-1999; 99US-0139899P.
PR 23-JUN-1999; 99US-0140353P.
PR 23-JUN-1999; 99US-0140354P.
PR 24-JUN-1999; 99US-0140695P.
PR 28-JUN-1999; 99US-0140823P.
PR 29-JUN-1999; 99US-0140991P.
PR 30-JUN-1999; 99US-0141287P.
PR 01-JUL-1999; 99US-0141842P.
PR 02-JUL-1999; 99US-0142154P.
PR 06-JUL-1999; 99US-0142055P.
PR 08-JUL-1999; 99US-0142390P.
PR 09-JUL-1999; 99US-0142920P.
PR 12-JUL-1999; 99US-0142977P.
PR 13-JUL-1999; 99US-0143542P.
PR 14-JUL-1999; 99US-0143622P.
PR 15-JUL-1999; 99US-0144005P.
PR 16-JUL-1999; 99US-0144085P.
PR 16-JUL-1999; 99US-0144086P.
PR 19-JUL-1999; 99US-0144325P.
PR 19-JUL-1999; 99US-0144331P.
PR 19-JUL-1999; 99US-0144332P.
PR 19-JUL-1999; 99US-0144333P.
PR 19-JUL-1999; 99US-0144334P.
PR 19-JUL-1999; 99US-0144335P.
PR 20-JUL-1999; 99US-0144352P.
PR 20-JUL-1999; 99US-0144632P.
PR 20-JUL-1999; 99US-0144884P.
PR 21-JUL-1999; 99US-0144814P.
PR 21-JUL-1999; 99US-0145086P.
PR 21-JUL-1999; 99US-0145088P.
PR 22-JUL-1999; 99US-0145085P.
PR 22-JUL-1999; 99US-0145087P.
PR 22-JUL-1999; 99US-0145089P.
PR 22-JUL-1999; 99US-0145192P.
PR 23-JUL-1999; 99US-0145145P.
PR 23-JUL-1999; 99US-0145218P.
PR 23-JUL-1999; 99US-0145224P.
PR 26-JUL-1999; 99US-0145276P.
PR 27-JUL-1999; 99US-0145918P.
PR 27-JUL-1999; 99US-0145919P.
PR 28-JUL-1999; 99US-0145951P.
PR 02-AUG-1999; 99US-0146386P.
PR 02-AUG-1999; 99US-0146388P.
PR 03-AUG-1999; 99US-0146389P.
PR 04-AUG-1999; 99US-0147038P.
PR 04-AUG-1999; 99US-0147204P.
PR 05-AUG-1999; 99US-0147302P.
PR 05-AUG-1999; 99US-0147192P.
PR 06-AUG-1999; 99US-0147260P.
PR 06-AUG-1999; 99US-0147303P.
PR 09-AUG-1999; 99US-0147416P.
PR 09-AUG-1999; 99US-0147493P.
PR 10-AUG-1999; 99US-0147935P.
PR 11-AUG-1999; 99US-0148171P.
PR 12-AUG-1999; 99US-0148319P.
PR 13-AUG-1999; 99US-0148341P.
PR 13-AUG-1999; 99US-0148565P.
PR 16-AUG-1999; 99US-0148684P.
PR 17-AUG-1999; 99US-0149368P.
PR 18-AUG-1999; 99US-0149175P.
PR 20-AUG-1999; 99US-0149426P.
PR 20-AUG-1999; 99US-0149722P.
PR 20-AUG-1999; 99US-0149723P.

PR 20-AUG-1999; 99US-0149929P.
PR 23-AUG-1999; 99US-0149902P.
PR 23-AUG-1999; 99US-0149930P.
PR 25-AUG-1999; 99US-0150566P.
PR 26-AUG-1999; 99US-0150884P.
PR 27-AUG-1999; 99US-0151065P.
PR 27-AUG-1999; 99US-0151066P.
PR 27-AUG-1999; 99US-0151080P.
PR 30-AUG-1999; 99US-0151303P.
PR 31-AUG-1999; 99US-0151438P.
PR 01-SEP-1999; 99US-0151930P.
PR 07-SEP-1999; 99US-0152363P.
PR 10-SEP-1999; 99US-0153707P.
PR 13-SEP-1999; 99US-0153758P.
PR 15-SEP-1999; 99US-0154018P.
PR 16-SEP-1999; 99US-0154039P.
PR 20-SEP-1999; 99US-0154779P.
PR 22-SEP-1999; 99US-0155139P.
PR 23-SEP-1999; 99US-0155486P.
PR 24-SEP-1999; 99US-0155488P.
PR 28-SEP-1999; 99US-0155659P.
PR 29-SEP-1999; 99US-0156458P.
PR 04-OCT-1999; 99US-0156596P.
PR 05-OCT-1999; 99US-0157117P.
PR 06-OCT-1999; 99US-0157553P.
PR 07-OCT-1999; 99US-0157865P.
PR 08-OCT-1999; 99US-0158029P.
PR 12-OCT-1999; 99US-0158232P.
PR 13-OCT-1999; 99US-0158369P.
PR 13-OCT-1999; 99US-0159293P.
PR 13-OCT-1999; 99US-0159294P.
PR 14-OCT-1999; 99US-0159295P.
PR 14-OCT-1999; 99US-0159328P.
PR 14-OCT-1999; 99US-0159330P.
PR 14-OCT-1999; 99US-0159331P.
PR 14-OCT-1999; 99US-0159637P.
PR 18-OCT-1999; 99US-0159638P.
PR 21-OCT-1999; 99US-0159584P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 22-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 25-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 26-OCT-1999; 99US-0161406P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 28-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 29-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:

Pred. No.: 229
Score: 8.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 1.20%
DB: 3
Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x AAG51289 (1-525)

QY 374 GTGGTTTCTGAGTTCTCTCGTT 397

Db 508 ValGlyLeuLeuSerSerLeuVal 515

RESULT 45
AAG22367

ID	AAg22387 standard; protein; 525 AA.	PR	18-JUN-1999;	99US-0139461P-
XX		PR	18-JUN-1999;	99US-0139462P-
AC	AAg22387;	PR	18-JUN-1999;	99US-0139463P-
XX		PR	18-JUN-1999;	99US-0139750P-
DT	17-OCT-2000 (first entry)	PR	18-JUN-1999;	99US-0139763P-
XX		PR	21-JUN-1999;	99US-0139817P-
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 25296.	PR	22-JUN-1999;	99US-0139899P-
XX		PR	23-JUN-1999;	99US-0140033P-
XX		PR	23-JUN-1999;	99US-0140034P-
KW	Protein identification; signal transduction pathway; metabolic pathway;	PR	24-JUN-1999;	99US-0140695P-
KW	hybridisation assay; genetic mapping; gene expression control; promoter;	PR	28-JUN-1999;	99US-0140823P-
XX	termination sequence.	PR	29-JUN-1999;	99US-0140991P-
OS	Arabidopsis thaliana.	PR	30-JUN-1999;	99US-0141287P-
XX		PR	01-JUL-1999;	99US-0141842P-
XX	EP1033405-A2.	PR	01-JUL-1999;	99US-0142154P-
XX		PR	02-JUL-1999;	99US-0142055P-
PD	06-SEP-2000.	PR	06-JUL-1999;	99US-0142390P-
XX		PR	08-JUL-1999;	99US-0142803P-
XX		PR	09-JUL-1999;	99US-0142920P-
PF	25-FEB-2000; 2000EP-00301439.	PR	12-JUL-1999;	99US-0142977P-
XX		PR	13-JUL-1999;	99US-0143542P-
XX		PR	14-JUL-1999;	99US-0143624P-
PR	05-MAR-1999;	PR	15-JUL-1999;	99US-0144005P-
PR	09-MAR-1999;	PR	16-JUL-1999;	99US-0144085P-
PR	23-MAR-1999;	PR	16-JUL-1999;	99US-0144086P-
PR	25-MAR-1999;	PR	19-JUL-1999;	99US-0144325P-
PR	29-MAR-1999;	PR	19-JUL-1999;	99US-0144331P-
PR	01-APR-1999;	PR	19-JUL-1999;	99US-0144332P-
PR	06-APR-1999;	PR	19-JUL-1999;	99US-0144333P-
PR	08-APR-1999;	PR	19-JUL-1999;	99US-0144334P-
PR	16-APR-1999;	PR	19-JUL-1999;	99US-0144335P-
PR	19-APR-1999;	PR	20-JUL-1999;	99US-0144352P-
PR	21-APR-1999;	PR	20-JUL-1999;	99US-0144632P-
PR	23-APR-1999;	PR	20-JUL-1999;	99US-0144684P-
PR	28-APR-1999;	PR	21-JUL-1999;	99US-0144814P-
PR	30-APR-1999;	PR	21-JUL-1999;	99US-0145086P-
PR	30-APR-1999;	PR	21-JUL-1999;	99US-0145088P-
PR	04-MAY-1999;	PR	22-JUL-1999;	99US-0145085P-
PR	05-MAY-1999;	PR	22-JUL-1999;	99US-0145087P-
PR	06-MAY-1999;	PR	22-JUL-1999;	99US-0145089P-
PR	07-MAY-1999;	PR	23-JUL-1999;	99US-0145192P-
PR	11-MAY-1999;	PR	23-JUL-1999;	99US-0145218P-
PR	14-MAY-1999;	PR	23-JUL-1999;	99US-0145224P-
PR	14-MAY-1999;	PR	26-JUL-1999;	99US-0145276P-
PR	14-MAY-1999;	PR	27-JUL-1999;	99US-0145913P-
PR	14-MAY-1999;	PR	27-JUL-1999;	99US-0145918P-
PR	18-MAY-1999;	PR	27-JUL-1999;	99US-0145919P-
PR	19-MAY-1999;	PR	28-JUL-1999;	99US-0145951P-
PR	20-MAY-1999;	PR	02-AUG-1999;	99US-0146386P-
PR	21-MAY-1999;	PR	02-AUG-1999;	99US-0146388P-
PR	24-MAY-1999;	PR	02-AUG-1999;	99US-0146389P-
PR	25-MAY-1999;	PR	03-AUG-1999;	99US-0147038P-
PR	27-MAY-1999;	PR	04-AUG-1999;	99US-0147204P-
PR	28-MAY-1999;	PR	04-AUG-1999;	99US-0147302P-
PR	01-JUN-1999;	PR	05-AUG-1999;	99US-0147792P-
PR	03-JUN-1999;	PR	06-AUG-1999;	99US-0147260P-
PR	04-JUN-1999;	PR	06-AUG-1999;	99US-0147303P-
PR	07-JUN-1999;	PR	06-AUG-1999;	99US-0147416P-
PR	08-JUN-1999;	PR	09-AUG-1999;	99US-0147493P-
PR	10-JUN-1999;	PR	09-AUG-1999;	99US-0147935P-
PR	10-JUN-1999;	PR	10-AUG-1999;	99US-0148171P-
PR	14-JUN-1999;	PR	11-AUG-1999;	99US-0148319P-
PR	16-JUN-1999;	PR	12-AUG-1999;	99US-0148341P-
PR	16-JUN-1999;	PR	13-AUG-1999;	99US-0148655P-
PR	17-JUN-1999;	PR	13-AUG-1999;	99US-0148684P-
PR	18-JUN-1999;	PR	16-AUG-1999;	99US-0149368P-
PR	18-JUN-1999;	PR	17-AUG-1999;	99US-0149175P-
PR	18-JUN-1999;	PR	18-AUG-1999;	99US-0149426P-
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149722P-
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149723P-
PR	18-JUN-1999;	PR	23-AUG-1999;	99US-0149929P-
PR	18-JUN-1999;	PR		99US-0149902P-

DB: 78 AsparGAlaAspSerArgSer 85
|||||
RESULT 47
ID AAR46679 standard; protein; 530 AA.
XX AAR46679;
AC AAR46679;
DT 25-MAR-2003 (revised)
DT 08-AUG-1994 (first entry)
XX
DE CD4-IgG2 chimeric heavy chain.
XX
KW CD4; gamma; heavy chain; chimeric; chimeric; immunoglobulin; HIV;
KW human immunodeficiency virus; radionuclide; toxin; therapy; treatment;
KW imaging; detection; targeting; immunoglobulin; IgG.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 1..204
FT /label= CD4 Region.
FT Region 205..302
FT /label= CH1 Region.
FT Region 303..314
FT /label= Hinge Region.
FT Region 315..423
FT /label= CH2 Region.
FT Region 424..530
FT /label= CH3 Region.
XX
PN MO9403191-A1.
XX
PD 17-FEB-1994.
XX
PP 06-AUG-1993; 93WO-US007422.
XX
PR 07-AUG-1992; 92US-00927931.
XX
PR (PROG-) PROGENICS PHARM INC.
XX
PI Alloway GP, Maddon PJ;
XX
DR WPI; 1994-065392/08.
DR N-PSDB; AA055751.
XX
PT Non-peptidyl toxin or radionuclide and CD4-gamma 2 or CD4-IgG2
PT immunoconjugates - used to kill HIV-infected cells and to image and
PT stage HIV infection.
XX
PS Disclosure; Fig 4; 142pp; English.
XX
CC A tetramer comprising 2 IgG2 heavy chains or two CD4-IgG2 chimeric heavy
CC chains and two kappa light chains or CD4-kappa light chains (AAR46680)
CC linked to a non-peptidyl toxin or a gamma radiation-emitting radionuclide
CC of low to moderate cytotoxicity. The resulting immunocytotoxic conjugate
CC the toxin can be used to kill HIV infected cells and to treat HIV
CC infected subjects to reduce the population of HIV infected cells. It can
CC also be used to reduce the likelihood of infection. The immunocytotoxic
CC comprising the radionuclide can be used to image HIV infected tissue, to
CC calculate the stage of HIV infection or the efficacy of an anti-HIV
CC treatment using the imaging technique and for determining the prognosis
CC of an HIV infected subject. (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 530 AA:
Alignment Scores:
Pred. No.: 229 Length: 530
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0

DB: 2 Gaps: 0
US-10-029-345A-108_COPY_532_2532 (1-2001) x AAR46679 (1-530)
QY 1819 GACAGAGCTGACTCGCGCGAGC 1842
DB 78 AsparGAlaAspSerArgSer 85
|||||
RESULT 48
ID AAY85080 standard; protein; 530 AA.
XX AAY85080;
AC AAY85080;
DT 19-JUN-2000 (first entry)
DT
XX
DE CD4-IgG2 chimeric heterotetramer heavy chain amino acid sequence.
XX
KW CD4-IgG2 chimeric heavy chain heterotetramer; immunocytotoxic; treatment;
KW cytotoxic radionuclide; cell surface glycoprotein; prevent; infection;
KW cellular immune response interaction mediator; HIV interaction; staging;
KW prognosis; envelope glycoprotein burden; human.
XX
OS Homo sapiens.
XX
PN US6034223-A.
XX
PD 07-MAR-2000.
XX
PP 07-JUN-1995; 95US-00477460.
XX
PR 07-AUG-1992; 92US-00927931.
PR 06-AUG-1993; 93WO-US007422.
PR 03-FEB-1995; 95US-00379516.
XX
PR (PROG-) PROGENICS PHARM INC.
XX
PI Alloway GP, Maddon PJ;
XX
DR WPI; 2000-269502/23.
DR N-PSDB; AA298856.
XX
PT New immunocytotoxic conjugate, used to treat, prevent or image human immune
PT deficiency virus infection, comprises radionuclide attached to
PT heterotetramer of CD4-immunoglobulin chimerae.
XX
PS Disclosure; Fig 4; 58pp; English.
XX
CC This sequence represents the CD4-IgG2 chimeric heavy chain amino acid
CC sequence from the CD4-IgG2 chimeric heterotetramer. The invention relates
CC to an immunocytotoxic conjugate comprising a cytotoxic radionuclide and a
CC heterotetramer of two heavy chains and two light chains. The cytotoxic
CC radionuclide is linked to either the heavy chains or the light chains, or
CC to all four chains, directly or through a bifunctional chelator. Both
CC heavy chains are chimeric CD4-Ig (immunoglobulin) G2 chains encoded by
CC vector CD4-IgG2HC-pRCMV (ATCC 75193) and both light chains are chimeric
CC CD4-kappa chains encoded by vector CD4-KLC-pRCMV (ATCC 75194). CD4 is a
CC non-polymorphic cell surface glycoprotein that is expressed on the
CC surface of helper T lymphocytes, cells of the monocyte/macrophage lineage
CC and dendritic cells. CD4 associates with major histocompatibility complex
CC (MHC) class II molecules on the surface of antigen presenting cells to
CC mediate efficient cellular immune response interactions. In humans CD4 is
CC the target of interaction with the human immunodeficiency virus HIV. The
CC immunocytotoxic conjugate is used to kill cells infected with HIV, and for treating
CC or preventing infection. It is also used for imaging HIV-infected tissues
CC (for staging or prognosis of infection, and for assessing efficacy of
CC treatments). The immunocytotoxic conjugate is also used to determine the HIV
CC envelope glycoprotein burden, once determined, this information is used
CC in the staging and prognosis of HIV infected patients. The
CC immunocytotoxic conjugate should be active against all strains of HIV (since the
CC CD4-SP120 interaction is essential for infection). The heterotetramers
CC are assembled intracellularly and secreted efficiently from mammalian
CC cells, allowing high recovery and purification from the culture medium.

CC They have longer half-life in serum and greater avidity than heavy chain
 CC dimers
 XX
 SQ Sequence 530 AA:

Alignment Scores:

Pred. No.:	229	Length:	530
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.20%	Indels:	0
DB:	3	Gaps:	0

US-10-029-345A-108_COPY_532_2532 (1-2001) x AAB85080 (1-530)

QY 1819 GACAGAGCTGACTCGGGCGGAGC 1842
 DB 78 AsparGAlaAspSerArgArgSer 85

RESULT 49
 AAB67323
 ID AAB67323 standard; protein; 530 AA.
 XX
 AC AAB67323;
 XX
 DT 23-APR-2001 (first entry)
 XX
 DE CD4-IgG2 chimeric heavy chain protein.
 XX
 KW Immunocjugate; chelator; chimeric; HIV; human immunodeficiency virus.
 KM
 OS Homo sapiens.
 XX
 PN US6177549-B1.
 XX
 PD 23-JAN-2001.
 XX
 PE 10-JUN-1999; 99US-00329916.
 XX
 PR 07-AUG-1992; 92US-00927931.
 PR 06-AUG-1993; 93WO-US007422.
 PR 03-FEB-1995; 95US-00379516.
 PR 07-JUN-1995; 95US-00477460.
 XX
 PA (PROG-) PROGENICS PHARM INC.
 XX
 PI Maddon PJ, Allaway GP;
 XX
 DR WPI; 2001-158582/16.
 XX
 PT Immunocjugate for treating human immunodeficiency virus-infected
 PT subject, consists of cytotoxic radionuclide linked to heterotetramer
 PT comprising two chimeric CD4-IgG2 heavy chains and two chimeric CD4-
 PT kappa/light chains.
 XX
 PS Disclosure; Fig 4; 43pp; English.
 XX
 CC The present invention relates to an immunocjugate, comprising a
 CC cytotoxic radionuclide linked, directly or via a bifunctional chelator,
 CC to a heterotetramer having two chimeric CD4-IgG2 heavy chains encoded by
 CC an expression vector CD4-IgG2HC-pRCMV and two chimeric CD4-kappa light
 CC chains encoded by an expression vector CD4-kLC-pRCMV. The invention is
 CC useful for killing human immunodeficiency virus (HIV)-infected cells, for
 CC the treatment and prevention of infection with HIV
 CC
 XX
 SQ Sequence 530 AA:

Alignment Scores:

Pred. No.:	229	Length:	530
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.20%	Indels:	0

DB: 4 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x AAB67323 (1-530)

QY 1819 GACAGAGCTGACTCGGGCGGAGC 1842
 DB 78 AsparGAlaAspSerArgArgSer 85

RESULT 50
 AAB80884
 ID AAB80884 standard; protein; 530 AA.
 XX
 AC AAB80884;
 XX
 DT 29-MAY-2001 (first entry)
 XX
 DE Human CD4-IgG2 chimeric heavy chain.
 XX
 KW Human; Anti-HIV; CD4-IgG2 chimeric heterotetramer;
 KM immunoglobulin gamma 2.
 XX
 OS Homo sapiens.
 XX
 PN US6187748-B1.
 XX
 PD 13-FEB-2001.
 XX
 PE 07-JUN-1995; 95US-00485372.
 XX
 PR 08-FEB-1991; 91US-00653684.
 PR 10-FEB-1992; 92WO-US001143.
 PR 08-DEC-1992; 92US-00960440.
 XX
 PA (PROG-) PROGENICS PHARM INC.
 XX
 PI Maddon PJ, Beaudry GA;
 XX
 DR WPI; 2001-264981/27.
 DR N-PSDB; AAF77830.
 XX
 PT Inhibiting human immunodeficiency virus (HIV) infection of a CD4+ cell,
 PT or treating a subject having CD4+ cells infected with HIV involves using
 PT CD4-IgG2 chimeric heterotetramer to form a complex with the HIV.
 XX
 PS Disclosure; Fig 4; 55pp; English.
 XX
 CC The present invention relates to a method for inhibiting infection of a
 CC CD4+ cell by HIV. The method comprises contacting the HIV with a CD4-IgG2
 CC chimeric heterotetramer to form a complex with the HIV (CD4 = cluster of
 CC differentiation 4; IgG2 = immunoglobulin gamma 2). CD4 is a cell surface
 CC glycoprotein that is expressed primarily on the surface of T cells. In
 CC man, CD4 is the target of interaction with HIV. The heterotetramer has
 CC two heavy and two light chains which are encoded by expression vectors
 CC CD4-IgG2HC-pRCMV (V1) and CD4-kLC-pRCMV (V2), respectively. The method
 CC is used to inhibit infection of a CD4+ cell by a HIV and to prevent CD4+
 CC cells of a subject from becoming infected with HIV. The method is also
 CC useful for treating a subject having CD4+ cells infected with HIV. The
 CC present sequence is human fusion protein: CD4-Ig2 chimeric heavy chain of
 CC the CD4-IgG2 chimeric heterotetramer. This sequence was used in the
 CC method of the present invention
 CC
 XX
 SQ Sequence 530 AA:

Alignment Scores:

Pred. No.:	229	Length:	530
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.20%	Indels:	0
DB:	4	Gaps:	0

US-10-029-345A-108_COPY_532_2532 (1-2001) x AAB80884 (1-530)

Qy 1819 GACGAGCTGACTCGCGCGGAGC 1842
|||
Db 78 AspArgAlaAspSerArgSer 85

Search completed: June 21, 2004, 13:09:22
Job time : 169.5 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 21, 2004, 13:09:27 ; Search time 93.5 Seconds

(without alignments)
12083.628 Million cell updates/sec

Title: US-10-029-345A-108_COPY_532_2532

Perfect score: 667
Sequence: 1 gttgcatcagccatgagat.....tggaaatcatcagagctcc 2001

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1163542 seqs, 282313646 residues

Word size: 1

Total number of hits satisfying chosen parameters: 349658

Minimum DB seq length: 473

Maximum DB seq length: 200000000

Post-processing: Listing first 75 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xld
-Q=/cgn2_1/USPTO_spool_p/US10029345/runat_21062004_124617_5758/app_query.fasta_1.2183
-DB=Published Applications_AA -QFMT=fasta -SUFFIX=oligo.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bites -START=1 -END=1 -MATRIX=oligo
-TRANS-human40.cdt -LIST=75 -DOCALLIGN=200 -THR_SCORE=quality -THR_MIN=1
-ALIGN=50 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=473
-MAXLEN=200000000 -USER=US10029345 @CGN_1_1_13 @runat_21062004_124617_5758
-NCPU=6 -ICPU=3 -NO_MMAP -LARGQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database: Published Applications_AA:*

1: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubppa/PCR_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubppa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubppa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubppa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubppa/PCRUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubppa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubppa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubppa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubppa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubppa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubppa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubppa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	ID	Description
--------	-------	-------	-------	--------	----	-------------

1	472	70.8	517	9	US-09-964-277-21	Sequence 21, Appl
2	472	70.8	665	9	US-09-816-494-2	Sequence 2, Appl1
3	472	70.8	665	9	US-09-964-277-2	Sequence 2, Appl1
4	472	70.8	665	12	US-10-072-012-680	Sequence 680, App
5	472	70.8	665	12	US-10-072-012-681	Sequence 681, App
6	472	70.8	665	12	US-10-168-506-14	Sequence 14, Appl
7	472	70.8	665	12	US-10-343-357-7	Sequence 7, Appl1
8	472	70.8	665	15	US-10-094-749-2312	Sequence 2312, Ap
9	472	70.8	665	15	US-10-377-027-26	Sequence 26, Appl
10	472	70.8	665	16	US-10-257-028-2	Sequence 2, Appl1
11	472	70.8	665	16	US-10-648-593-240	Sequence 240, App
12	472	70.8	665	16	US-10-648-593-247	Sequence 247, App
13	472	70.8	665	12	US-10-072-012-679	Sequence 679, App
14	472	70.8	665	12	US-10-072-012-703	Sequence 703, App
15	472	70.8	665	12	US-10-425-114-54204	Sequence 54204, A
16	472	70.8	665	12	US-10-072-012-258	Sequence 258, App
17	394	59.1	660	12	US-10-072-012-256	Sequence 256, App
18	258	38.7	660	12	US-10-296-115-1259	Sequence 1259, Ap
19	67	10.0	660	12	US-10-072-012-682	Sequence 682, App
20	64	9.6	660	12	US-10-072-012-683	Sequence 683, App
21	20	3.0	501	12	US-10-072-012-702	Sequence 702, App
22	20	3.0	625	12	US-10-072-012-659	Sequence 659, App
23	20	3.0	663	12	US-10-072-012-700	Sequence 700, App
24	12	1.8	616	12	US-10-072-012-266	Sequence 266, App
25	10	1.6	498	12	US-10-282-122A-57268	Sequence 57268, A
26	8	1.2	503	14	US-10-225-060-11	Sequence 11, Appl
27	8	1.2	509	12	US-10-214-812-2	Sequence 2, Appl1
28	8	1.2	509	12	US-10-214-812-3	Sequence 3, Appl1
29	8	1.2	509	14	US-10-215-224-2	Sequence 2, Appl1
30	8	1.2	509	14	US-10-215-224-3	Sequence 3, Appl1
31	8	1.2	509	14	US-10-225-060-10	Sequence 10, Appl
32	8	1.2	513	9	US-09-833-745-57	Sequence 57, Appl
33	8	1.2	529	12	US-10-424-599-231545	Sequence 231545,
34	8	1.2	530	8	US-08-485-163-5	Sequence 5, Appl1
35	8	1.2	530	9	US-09-766-995-4	Sequence 4, Appl1
36	8	1.2	532	10	US-09-939-537-6	Sequence 6, Appl1
37	8	1.2	532	11	US-09-243-008-6	Sequence 11, Appl
38	8	1.2	534	14	US-10-046-231-22	Sequence 22, Appl
39	8	1.2	544	14	US-10-156-761-11387	Sequence 11387, A
40	8	1.2	545	16	US-10-628-088-415	Sequence 415, App
41	8	1.2	555	10	US-09-934-008-4	Sequence 4, Appl1
42	8	1.2	575	11	US-09-243-008-4	Sequence 4, Appl1
43	8	1.2	590	9	US-09-934-060A-13	Sequence 13, Appl
44	8	1.2	653	14	US-10-062-937B-5	Sequence 5, Appl1
45	8	1.2	653	15	US-10-391-399-73	Sequence 73, Appl
46	8	1.2	677	9	US-09-833-745-39	Sequence 39, Appl
47	8	1.2	677	15	US-10-369-493-6901	Sequence 6901, Ap
48	8	1.2	711	9	US-09-792-630-33	Sequence 33, Appl
49	8	1.2	711	10	US-09-953-351-33	Sequence 33, Appl
50	8	1.2	711	13	US-10-080-376-33	Sequence 33, Appl
51	8	1.2	711	13	US-10-082-671-33	Sequence 33, Appl
52	8	1.2	711	14	US-10-097-100-33	Sequence 33, Appl
53	8	1.2	711	14	US-10-023-208-33	Sequence 23, Appl
54	8	1.2	720	9	US-09-934-060A-2	Sequence 2, Appl1
55	8	1.2	720	9	US-09-934-060A-4	Sequence 4, Appl1
56	8	1.2	774	12	US-10-424-599-257796	Sequence 257796,
57	8	1.2	788	14	US-10-073-118-26	Sequence 26, Appl
58	8	1.2	912	9	US-09-987-482-2	Sequence 2, Appl1
59	8	1.2	1222	12	US-10-282-122A-69799	Sequence 69799, A
60	8	1.2	1404	12	US-10-425-114-55163	Sequence 55163, A
61	8	1.2	2630	15	US-10-369-493-3693	Sequence 3693, Ap
62	8	1.2	2883	8	US-08-681-219-32	Sequence 32, Appl
63	8	1.2	2883	9	US-09-987-482-1	Sequence 1, Appl1
64	8	1.2	2883	11	US-09-230-111C-30	Sequence 30, Appl
65	8	1.2	2883	12	US-10-392-113-21	Sequence 21, Appl
66	8	1.2	2883	14	US-10-092-138-30	Sequence 30, Appl
67	8	1.2	2883	16	US-10-408-765A-1970	Sequence 1970, Ap
68	8	1.2	2884	12	US-10-267-502-370	Sequence 370, App
69	8	1.2	2884	12	US-10-267-502-372	Sequence 372, App
70	8	1.2	3972	14	US-10-156-761-8476	Sequence 8476, Ap
71	8	1.2	3972	16	US-10-204-862A-4	Sequence 4, Appl1
72	8	1.2	3972	16	US-10-204-862A-8	Sequence 8, Appl1
73	8	1.2	5532	14	US-10-156-761-8480	Sequence 8480, Ap

	US-10-204-862A-7	US-10-282-122A-52396	Sequence 7, Appl. 52396, A
C 74	16	12	Sequence 7, Appl. 52396, A
C 75	1.2	1.1	Sequence 52396, A

ALIGNMENTS

RESULT 1
US-09-964-277-21

```

? Sequence 21, Application US/09964277
? Patent No. US20020137170A1
? GENERAL INFORMATION:
? APPLICANT: Lucche, Ralf M.
? APPLICANT: Wei, Bo
? TITLE OF INVENTION: DSP-16 DUAL-SPECIFICITY PHOSPHATASE
? FILE REFERENCE: 200125.434
? CURRENT APPLICATION NUMBER: US/09/964.277
? CURRENT FILING DATE: 2001-09-25
? NUMBER OF SEQ ID NOS: 22
? SOFTWARE: FaastSD for Windows Version 4.0
? SEQ ID NO 21
? LENGTH: 517
? TYPE: PRF
? ORGANISM: Homo sapiens
US-09-964-277-21

```

Alignment Scores:

Pred. No.:	0	517
Score:	472.00	Matches: 472
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	70.76%	Indels: 0
DB:	9	Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-09-964-277-21 (1-517)

QY	586	ACCGTCCAAAGCCGTGACTTATCCCGGAGTCTCAATTTCCGTCGTCGTCGAATGAC	6145
Db	46	Thrcysprolysproaephelleproglusehishpheleuagvalprovalaasasp	65
QY	646	AGCTTTGTGAGAAAATTTTGCCTGTGGACAATACTAGTAGATTTCATTGAGAAAGCA	705
Db	66	Sertheysgsluwylsileuprotplleaumaplyservalaeppheilegluluala	85
QY	706	AAAGCCTCCATGGATGTGTTCAGAGCACTGTTCAGTCGGGAGTCCCGTCGGCCACC	765
Db	86	Lyslaaerxsnnglycysvalleuvalhiscysleuilaaglylileseraagseralath	105
QY	766	ATGCGTATCGCCTPACATCATGAAGAGATGACATGTCTTTAGATGAAGCTTTACAGATT	825
Db	106	lleallealleatylrilemetlysarwecsaepmetserleuamglnualatylrathgrphe	125
QY	826	GTGAAGAAGAAAAGACCTCATATCTCCAACTTCATTTTCTGGGCAACTCTCGAC	885
Db	126	Vallysgulysarkprothrillileserproaenpheanpheuonglylnleuleaarp	145
QY	886	TATGAAGAAGATTAAAGACCAAGACTGAGGATCAGAGGCAAGAACTCAAGCTG	945
Db	146	Tyglulwelysillelysasnghlnhrglylaaerxgylprolysserlylseulyseu	165
QY	946	CTGCACTGGAGAAAGCCAAATGAACCTGTCCCTGTCTTCAGAGGGGTGACAGAAAAGC	1004
Db	166	lenhileuendulysproasnghluprovalaalaalserglnuglylnlysser	185
QY	1006	GAGACGCCCTCAGTCCACCTGTGSCCAACTCTGTACCTCAGAGGAGAGAGCAAAAG	1065
Db	186	glulthrproleuserproprocysalaaspherlaathrserglnualaalaaglylnaag	205
QY	1066	CCCGTCATCCCGCCAGCGGTGCCCAAGCTGCCCAAGCTGAGAGCGCTGTGTAGAGAC	1122
Db	206	Provalinhiproalaservalaproservalaproservalglnproserleuenglnabp	225
QY	1126	AGCGCGCTGGTACAGCGCTCAGTGGGCTGCACTGTCCGACACAGGCTGAAAGACAC	1185

Db 226 SerProLeuValGlnAlaLeuSerGlyLeuHisLeuSerAlaAspArgLeuGlnAspSer 245
 QY 1186 AATTAGCTCAACCGCTTCCTCTCTCGATATCAAAATCATGTTTCATATTACGCCAGCATG 1245
 Db 246 AsnIleuSLeuYfARsSerPheSerLeuAspIleuYSerValSerTYrSerAlaSerMet 265
 QY 1246 GCAGCATCCTTACATGAGCTTCTCCATCAGAAAGATGCTTTGGAAATCAAAACCTTCC 1305
 Db 266 AlaIaSerLeuHisGlyPheSerSerSerGlnuAspAlaLeuGlnuTYrLYrIlyProSer 285
 QY 1306 ACTTACTCTGGATGGGACCAACAAGCTTATGCCAGTTCTTCCTCTGTTACAGAACTATCGAG 1365
 Db 286 ThrThrLeuAspGlyThrAsnLYsLeuGlySerGlnPheSerProValGlnGlnLeuSerGln 305
 QY 1366 CAGACTCCCGAAACCAAGTCTGATATAGGAGAGACCGACAGATCCCGCAAGAGCTGCACACC 1425
 Db 306 GlnThrProGlnuThrSerProAspLYsGlnuAlaSerIleProLYsLYsLeuGlnuThr 325
 QY 1426 GCCAGCGCTTCAACAGCCAGACCAAGCGATTGCGTCAAGAACCAAGACAGCATGCGC 1485
 Db 326 AlaArgProSerAspSerGlnSerLYsArgLeuHisSerValArgThrSerSerSerGly 345
 QY 1486 ACCGCGCCAGAGATCCCTTTATCTTCCACTGATGATGAGTGGAGCGCTGGAGCAATTAC 1545
 Db 346 ThrIaGlnHisSerLeuSerProLeuHisArgSerGlySerValGlnuAspAsnTYr 365
 QY 1546 CACACCAAGCTTCCTTTTCGGCTTTCACAGACCCAGACCACTTCAGAACTGCTGCGC 1605
 Db 366 HisThrSerPheLeuPheGlyLeuSerThrSerGlnGlnHisLeuThrLYsSerAlaGly 385
 QY 1606 CTGGGCGCTTAAGGCGTGGCACTCGGAAATCTTGCGCCCGACAGACCTTACCCCTTCCCTG 1665
 Db 386 LeuGlyLeuLYsGlyTYrPheAspSerAspIleLeuAlaProGlnuThrSerThrProSerLeu 405
 QY 1666 ACCAGCAGCTGGTATTTTGGCCACAGAGTCTCTCACACTTACTCTGCTCAGCCATCTAC 1725
 Db 406 ThrSerSerTrpTYrPheAlaThrGlnuSerSerHisPheTYrSerAlaSerAlaIleTYr 425
 QY 1726 GGAGGCGTGGCAAGTACTCTGCTTACAGCTGCAGCCAGCTGCGCCACTTGGGAGCAAA 1785
 Db 426 GlyLYsSerAlaSerTYrSerAlaTYrSerCysSerGlnLeuProThrCysGlyAspGln 445
 QY 1786 GTCAATTCGTGGCGAGCGGCGAGCAAGGCAAGACAGAGCTGACTCGGCGGAGACTGG 1845
 Db 446 ValTYrSerValArgArgArgGlnuYfProSerAspArgAlaAspSerArgArgSerTrp 465
 QY 1846 CATGAAGAGAGCCCTTTGAAAAAGCATTTAAACGAGAAAGCTGCCAAATGCAATTTGGA 1905
 Db 466 HisGlnGlnuSerProPheGlnuYfGlnPheLYsArgArgSerCysGlnuMetGlnuPheGly 485
 QY 1906 GAGAGCATCATTCACAGAAACAGGTCAACGGAAGACTTGGGAAAGTGGGCGTCACTCT 1965
 Db 486 GlnSerIleuSerGlnuAsnArgSerArgGlnuLeuGlyLYsValGlySerGlnSer 505
 QY 1966 AGCTTTTCGGGCGAGCATGAAATCATTTGAGTCTCC 2001
 Db 506 SerPheSerClySerMetGlnuIleGlnuValSer 517
 RESULT 2
 US-09-816-494-2
 Sequence 2, Application US/09816494
 Patent No. US20020034807A1
 GENERAL INFORMATION:
 APPLICANT: Meyers, Rachel A.
 TITLE OF INVENTION: 38692 AND 21117, NOVEL DUAL SPECIFICITY
 TITLE OF INVENTION: PHOSPHATASE MOLECULES AND USES THEREFOR
 FILE REFERENCE: 10448-030002
 CURRENT APPLICATION NUMBER: US/09/816,494
 CURRENT FILING DATE: 2001-03-23
 PRIOR APPLICATION NUMBER: US 60/191,858
 PRIOR FILING DATE: 2000-03-24
 NUMBER OF SEQ ID NOS: 10
 SOFTWARE: PaeSeq for Windows Version 4.0

```
; SEQ ID NO 2
; LENGTH: 665
; TYPE: prf
; ORGANISM: Homo sapiens
US-09-816-494-2

Alignment Scores:
Pred. No.: 0          Length: 665
Score: 472.00        Matches: 472
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 70.76%          Indels: 0
DB: 9                  Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x US-09-816-494-2 (1-665)

QY 586 ACCGTGCAAGAGCTGACTTTATCCCGAGTCTCATTTCTCGTGCCTGTGAATGAC 645
   |||
Db 194 ThrCysProLysProAspPheIleProGluSerHisPheLeuAlaGValProValAsnAsp 213

QY 646 AGCTTTGTGAGAAAATTTTCCGCGTGTGGACAATCGATGATTCTTGAAGAACA 705
   |||
Db 214 SerPheCysGluLysIleLeuProTyrLeuAspLysSerValAspPheIleGluLysAla 233

QY 706 AAGAGCTCAATGATGTCTTCACTGACAGTCTGTTAGCTGGGACTCCCGCTCCGACAC 765
   |||
Db 234 LysAlaSerAsnGlyCysValLeuValHisCysLeuAlaGlyIleSerArgSerAlaThr 253

QY 766 ATCGTATCGCTACATCATGAAGAAGATGACATGCTTTAGATGAGCTTACAGATT 825
   |||
Db 254 IleAlaIleAlaIleTyrIleMetLysArgMetAspMetSerLeuAspGluAlaTyrArgPhe 273

QY 826 GTGAAAGAAAAGACCTTACTATATCTCAAACTTAATTTCTGGGCCAACTCTTGAC 885
   |||
Db 274 ValLysGluLysArgProThrIleSerProAsnPheAsnPheLeuGlyGlnLeuAsp 293

QY 886 TTGTGAGAGAGATTAAAGACGATGAGAGATCAAGGCGCAAGAGCAAACTCAAGCTG 945
   |||
Db 294 TyrGluLysLysIleLysAsnGlnThrGlyAlaSerGlyProLysSerLysLeuLysLeu 313

QY 946 CTGCACTTGAGAGAGCAATGAACTGCTCCCTGCTGTCTCAGAGAGGTGAGAGAAAGC 1005
   |||
Db 314 LeuHisLeuGluLysProAsnGlnProValProAlaValSerGluGlyGlnLysSer 333

QY 1006 GAGAGCGCCCTCAGTCACTCTGTGCGCACTGTCTAAGCTCAGAGCGACGAGCAAGG 1065
   |||
Db 334 GluThrProLysSerProProCysAlaAspSerAlaThrSerGluAlaGlyGlnArg 353

QY 1066 CCGCGCAATCCGCGCGAGCGTCCGAGCGTCCAGCGTCCGCGCTGTGAAGAGAC 1125
   |||
Db 354 ProValHisProAlaSerValProSerValProSerValGlnProSerLeuLeuAsp 373

QY 1126 ACCCGCGTGTGACAGCGCTCAGTGGGCTGACCTTCCGACAGACAGCTGGAAGACAGC 1185
   |||
Db 374 SerProLeuValGlnAlaLeuSerGlyLeuHisLeuSerAlaAspArgLeuGluAspSer 393

QY 1186 AATTAAGCTCAAGGTTCTCTCTCTGTGATATCAAACTCAATTCATATTCAAGCCAGCATG 1245
   |||
Db 394 AsnLysLeuLysArgSerPheSerLeuAspIleLysSerValSerTyrSerAlaSerMet 413

QY 1246 GGAGCATCTTACATGGCTTCCCTCATGAGAAGAGCTTGGAAATCAACAACCTTCC 1305
   |||
Db 414 AlaAlaSerLeuHisGlyPheSerSerSerGluAspAlaLeuGlnLysTyrLysProSer 433

QY 1306 ACTACTCTGGATGGAGCAACAAGCTATGCAAGTTCTCCCTGTTGAGAACTATGAGAG 1365
   |||
Db 434 ThrThrLeuAspGlyThrAsnLysLeuCysGlnPheSerProValGlnGlnLeuSerGlu 453

QY 1366 CAGACTCTCCGAAAACAAGTCTCTGTAAGAAGAAAGCAAGATCCCAAGAGCTGACAGCC 1425
   |||
Db 454 GlnThrProGluThrSerProAspLysGluGluAlaSerIleProLysLysLeuGlnThr 473

QY 1426 GCACAGCCCTTCAGACAGCCAGCAAGCAAGCATTCGTTCCGTCAGAACCAAGCAAGCTGAC 1485
```

```
Db 474 AlaArgProSerAspSerGlnSerLysArgLeuHisSerValArgThrSerSerGly 493

QY 1486 ACCGCCAGAGAGTCCCTTTATCTCCACTGCATCGAAGTGGAGCGGTGAGACAAATTAC 1545
   |||
Db 494 ThrAlaGlnArgSerLeuLeuSerProLeuHisArgSerGlySerValGluAspAsnTyr 513

QY 1546 CACACAGCTTCCTTTTGGCGCTTCCACAGCCAGAGCAAGCTCAGCAAGAGCTGCTGAC 1605
   |||
Db 514 HisThrSerPheLeuPheGlyLeuSerThrSerGlnGlnHisLeuThrLysSerAlaGly 533

QY 1606 CTGGGCTTAAAGGGGTGGCACTCGGATATCTTGGCCCCCAGACCTTAACCCCTTCCCTG 1665
   |||
Db 534 LeuGlyLeuLysGlyTyrHisSerAspIleLeuAlaProGlnThrSerThrProSerLeu 553

QY 1666 ACCAGCAGCTGGTATTTTCCACAGAGTCTTCAACTTACTGTGCTCAGCCATCTAC 1725
   |||
Db 554 ThrSerSerTyrPheAlaThrGluSerSerHisPheTyrSerAlaSerAlaIleTyr 573

QY 1726 GGAGGCAAGTCCAGTTACTGCGCTTACAGCTGACAGCTGACCTGCGGAGAGCA 1785
   |||
Db 574 GlyLysSerAlaSerTyrSerAlaTyrSerCysSerGlnLeuProThrCysGlyAspGln 593

QY 1786 GCTATTCGTGTCGAGCGCGAGAGCAAGCTGACAGAGCTGACTCCGCGGAGAGCTGG 1845
   |||
Db 594 ValTyrSerValArgArgArgGlnLysProSerAspArgAlaAspSerArgArgSerTyr 613

QY 1846 CATGAAGAGAGCCCTTTGAAAAGCATTTAAAGCCAGAAAGCTGCCAAATGGAATTGGA 1905
   |||
Db 614 HisGluGluSerProPheGluLysGlnPheLysArgArgSerCysGlnMetGluPheGly 633

QY 1906 GAGACATCATATTCAGAAACAGGTCACGGGAAAGCTGGGAAATGGGCACTAGTCT 1965
   |||
Db 634 GluSerIleMetSerGluAsnArgSerArgGluGluLeuGlyLysValGlySerGlnSer 653

QY 1966 AGCTTTTCGGGAGCATGGAATTCATTGAGCTCTCC 2001
   |||
Db 654 SerPheSerGlySerMetGluIleIleGluValSer 665

RESULT 3
US-09-964-277-2
; Sequence 2, Application US//09964277
; Patent No. US20020137170A1
; GENERAL INFORMATION:
; APPLICANT: Lucne, Ralf M.
; APPLICANT: Wei, Bo
; TITLE OF INVENTION: DSP-16 DUAL-SPECIFICITY PHOSPHATASE
; FILE REFERENCE: 200125_434
; CURRENT APPLICATION NUMBER: US/09/964,277
; CURRENT FILING DATE: 2001-09-25
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 665
; TYPE: prf
; ORGANISM: Homo sapiens
US-09-964-277-2

Alignment Scores:
Pred. No.: 0          Length: 665
Score: 472.00        Matches: 472
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 70.76%          Indels: 0
DB: 9                  Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x US-09-964-277-2 (1-665)

QY 586 ACCGTGCAAGAGCTGACTTTATCCCGAGTCTCATTTCTCGTGCCTGTGAATGAC 645
   |||
Db 194 ThrCysProLysProAspPheIleProGluSerHisPheLeuAlaGValProValAsnAsp 213

QY 646 AGCTTTGTGAGAAAATTTTCCGCGTGTGGACAATCGATGATTCTTGAAGAACA 705
```

Db 214 SerPheCysGluIuylIleuProTrpIeuAspIysSerValAspPheIleGluIuylAspAla 233
QY 706 AAAGCCTCCAAATGATGTTCTTACTGACCTTTAGCTGGGATCTCCGCTCCGCCAAC 765
Db 234 LysAlaSerAsnGlyCysValIleuValHisCysLeuAlaGlyIleSerArgSerAlaThr 253
QY 766 ATGCTATCGCCTATCATGATGATGAGATGACATGCTCTTATAGTAAAGCTTACAGATT 825
Db 234 IleAlaIleAlaIleAlaIleIleMetLysArgMetAspMetSerIleuAspGluAlaIleArgPhe 273
QY 826 GTGAAGAAAAAGACCTTACTATATCTCCAACTTCAATTTTCTGGGCCCACTCTCGAC 885
Db 274 ValIysGluIuylsArgProThrIleSerProAsnPheAsnPheLeuGlyIleuLeuAsp 293
QY 886 TATGAGAAGAAATTAAGAACACAGCTGAGACATCAGGGGCCCAAGACCAACTCAAGCTG 945
Db 294 TyrGluIuylsIleLysAsnGlnThrGlyAlaSerGlyProLysSerLysLeuLysLeu 313
QY 946 CTGCACTGGAAGAACCAATGAACTGTCCCTGCTGTCTCAGAGGGTGGACAGAAAGC 1005
Db 314 LeuHisLeuGluIuylsProAsnGluProValProAlaValSerGluGlyIleLysSer 333
QY 1006 GAGAGCCCTCCATCCACCTGTCGCGACTCTGCTCTTCAAGAGGACAGACAAAGG 1065
Db 334 GluHisProLeuSerProProCysAlaAspSerAlaThrSerGluAlaIleGlyIleArg 353
QY 1066 CCGGTGATCCCGCCAGCAGTCCAGCGCTGCCAGCGTCCGCTGTTAAGAGGAC 1125
Db 354 ProValHisProAlaSerValProSerValProSerValGlnProSerLeuLeuGluAsp 373
QY 1126 AGCCCGCTGTGATACAGGGGCTCAGTGGGCTGCACCTGTCCGACAGACAGCTGAAGACAC 1185
Db 374 SerProLeuValGlnAlaLeuSerGlyLeuHisIleLeuSerAlaAspArgLeuGluAspSer 393
QY 1186 AATAAGCTCAAGGCTTCTCTCTCTGATATCAATTCAGTTTCATATTCAGCCAGCAG 1245
Db 394 AsnLysLeuLysArgSerPheSerIleuAspIleLysSerValSerLysSerIleAspMet 413
QY 1246 GCAGCATCTTAAATGAGCTTCTCTCATCAGAAAGTCTTGGAACTTACAAACCTTCC 1305
Db 414 AlaAlaSerLeuHisGlyPheSerSerSerGluAspAlaLeuGluIuylTrpIysProSer 433
QY 1306 ACTACTGTGATGGAGCAACAAGACATAGCCAGTTCTCCCTGTTCAAGAACTATGAGAG 1365
Db 434 ThrThrLeuAspIleLysThrAsnLysLeuCysGlnPheSerProValGlnIleuLeuSerGln 453
QY 1366 CAGACTCCCGAAACAGCTCTGATAGAGAGAACAGACATCCCAAGAAAGCTGAGAGCC 1425
Db 454 GlnThrProGluThrSerProAspLysGluGluAlaSerIleProLysLysLeuGlnThr 473
QY 1426 GCCAGGCTTTCAGACAGCCAGCAAGCAAGATTGCAATTCGATCAGAACAGCAGCAGTGGC 1485
Db 474 AlaArgProSerAspSerIleSerLysArgLeuHisIleSerValArgThrSerSerIle 493
QY 1486 ACCGCCCAAGAGGCTCTTATCTCCACATGATCCAGAGTGGAGAGCTGAGAGCAATTAC 1545
Db 494 ThrAlaGlnArgSerIleuLeuSerProLeuHisArgSerGlySerValGlnAspAsnTrp 513
QY 1546 CACACAGCTTCTTTTGGCCTTTCCACAGCAGCAGACCTCAGCAAGCTGCTGGC 1605
Db 514 HisThrSerPheLeuPheGlyLeuSerThrSerGlnGlnHisIleuThrLysSerAlaGly 533
QY 1606 CTGGGCTTAAAGGCTGGGACCTCGCATCTTTGGCCCCCAGACCTTACCTTCCCTG 1665
Db 534 LeuGlyLeuLysGlyTrpHisIleSerAspIleLeuAlaProGlnThrSerThrProSerLeu 553
QY 1666 ACCACAGCTGATATTTGGCAGAGCTCCTCAACTCTACTGCTCGCTGAGCATCTAC 1725
Db 554 ThrSerSerTrpLysPheAlaThrGluSerSerHisPheTrpSerAlaSerAlaIleTrp 573
QY 1726 GAGGCACTGCCAGTTACTTGTGCTTACAGCTGACGCCAGCTGCCCATTTGGAGACCAA 1785

Db 574 GlyIylSerAlaSerTrpSerAlaTrpSerCysSerGlnLeuProThrCysGlyAspGln 593
QY 1786 GTCTATTCTGTGGCGAGCGGACGAAGCCAAATGACAGAGCTGACTCGGCGGAGCTGG 1845
Db 594 ValTrpSerValArgArgArgGlnLysProSerAspArgAlaAspSerArgArgSerTrp 613
QY 1846 CATGAAGAGACCCCTTTGAAAAAGCATTTAAACGCAAGCTGCCAAATGGAATTGGA 1905
Db 614 HisGluGluSerProPheGluIuylsGlnPheLysArgArgSerCysGlnMetGluPheGly 633
QY 1906 GAGAGCATCATCTCAGAAACAGCTCAGCGGAGAGCTGGGGAAGTGGGAGTCACTCT 1965
Db 634 GluSerIleMetSerGluAsnArgSerArgGlnGluLeuGlyValGlySerGlnSer 653
QY 1966 ACCTTTTGGCGAGCATGGAATCATTTGAGGCTTCC 2001
Db 654 SerPheSerGlySerMetGluIleIleGluValSer 665
RESULT 4
US-10-072-012-680
Sequence 680, Application US/10072012
Publication No. US20040033493A1
GENERAL INFORMATION:
APPLICANT: Tchernov, Velizar
APPLICANT: Spytek, Kimberly
APPLICANT: Zethusen, Bryan
APPLICANT: Patnrajan, Meera
APPLICANT: Shinkets, Richard
APPLICANT: Li, Li
APPLICANT: Gangolli, Esna
APPLICANT: Padiganu, Muralidhara
APPLICANT: Anderson, David W.
APPLICANT: Rastelli, Luca
APPLICANT: Miller, Charles E.
APPLICANT: Gerlach, Valerie
APPLICANT: Taupier Jr, Raymond J.
APPLICANT: Gusev, Vladimir Y.
APPLICANT: Colman, Steven D.
APPLICANT: Wolenc, Adam R.
APPLICANT: Pena, Carol E. A
APPLICANT: Furtak, Katarzyna
APPLICANT: Grose, William M.
APPLICANT: Alsobrook II, John P.
APPLICANT: Lepley, Denise M.
APPLICANT: Rieger, Daniel K.
APPLICANT: Burgess, Catherine E.
TITLE OR INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-258
CURRENT APPLICATION NUMBER: US/10/072, 012
PRIOR FILING DATE: 2002-01-31
PRIOR APPLICATION NUMBER: 60/265,102
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: 60/265,514
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,517
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,412
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,395
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/266,406
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 60/266,767
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: 60/267,057
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/266,975
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/267,459
PRIOR FILING DATE: 2001-02-08
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 1391
SOFTWARE: PatentIn Ver. 2.1

```
/ SEQ ID NO 680
/ LENGTH: 665
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-072-012-680

Alignment Scores:
Pred. No.: 0          Length: 665
Score: 472.00        Matches: 472
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 70.76%  Indels: 0
DB: 12          Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x US-10-072-012-680 (1-665)

QY 586 ACCCTGCAAAAGCTGACTTTATCCCGAGTCTCATTTCTGCGTGTGCTGGAATGAC 645
DB 194 ThrCybProluSerProAmphe1leProGluSerHisPheLeuAlyValProValaAmp 213
QY 646 AGCTTTTGAGAAAATTTTCCCGTGTGAGCAATTCAGTAGATTTCATTGAGAAACA 705
DB 214 SerPheCygluLyse1leuProTrpLeuAplySerValAasphe1legluLybAla 233
QY 706 AAGGCTCAATGAGTGTCTTCAAGCACTGTTAGCTGGGATCTCCGCTCCGACAC 765
DB 234 LybAlaSerAmp1CyvalLeuValHisCybLeuAlaGly1leSerAgsSerAlaTrp 253
QY 766 ATGCTATGCGCTACATCATGAGAGAGATGACATGCTTTAGATGAGCTTACAGATT 825
DB 254 1leAla1leAla1leTr1leMetLybAgsMetAaspMetSerLeuAplyAlaTrpAgsPhe 273
QY 826 GTGAAAGAAAAGACCTTACTATATCTCCAACTTCAATTTTGGGCGCACTCTGAGAC 885
DB 274 ValLybGluLybAgsProThrl1leSerProAmpAaspPheLeuGluLeuAmp 293
QY 886 TATGAGAGAAGATTAGAACCAGATGAGCATCAGGGCGCAAGAGCAAACTCAGACG 945
DB 294 TyrgluLybLyse1leuAmpGluTrgLyAlaSerGlyProLySerLybLeuLybLeu 313
QY 946 CTGCACTGAGAGAGCAATGAACTGTCCCTGCTGTCTCAAGAGGTGAGCAAGAAAC 1005
DB 314 LeuHisLeuGluLybProAmpGluProValProAlaValSerGluGlyGluLybSer 333
QY 1006 GAGAGCCCTCATGCTCACCCTGTGCGACTGTGCTAAGCTCAGAGCGACAGCAAG 1065
DB 334 GluTrpProluSerProProCybAlaApsSerAlaThrSerGluAlaGlyGlnAgs 353
QY 1066 CCGGTCATCCGCGCAGCGGCCAGCGGCCAGGGGTGAGCGCGTGGCTTGAAGAGAC 1125
DB 354 ProValHisProAlaSerValProSerValProSerValGlnProSerLeuGluAmp 373
QY 1126 AGCCCGCTGATCAGAGCGCTCAGTGGGCTGACCTTGCCGACAGACAGCTGAGACAC 1185
DB 374 SerProLeuValGlnAlaLeuSerGlyLeuHisLeuSerAlaApsArgLeuGluApsSer 393
QY 1186 AATAAGCTCAAGCGTTCTTCTCTGTGATATCAAAATCAAGTTTCATATTACGCACATG 1245
DB 394 AenLybLeuLybAgsSerPheSerLeuAply1leLySerValSerTySerAlaSerMet 413
QY 1246 GAGAGCTCTTCAATGAGCTTCTCCCTCATGAGAGAGCTTTGGAATACAAACCTTCC 1305
DB 414 AlaAlaSerLeuHis1leGlyPheSerSerSerGlnApsAlaLeuGluTrpLybProSer 433
QY 1306 ACTACTCTGGATGAGCAACAAGCTATGCGCACTTCCCTGTTGAGAACTATCGAGAG 1365
DB 434 ThrThrLeuApsGlyThrAenLybLeuCybGlnPheSerProValGlnGluLeuSerGlu 453
QY 1366 CAGACTCCCGAAAACAAGCTCTGATTAAGAGAGAAACCAAGCATCCCAAGAACTGACAGC 1425
DB 454 GlnThrProGluTrpThrSerProAmpLybGluGluAlaSer1leProLybLybLeuGlnThr 473
QY 1426 GCCAGGCTTCAGACAGCCAGCAAGCATTTGCAATTCGCTCAAGAACAGCAAGCATGGCC 1485
```

```
DB 474 AlaArgProSerAaspSerGlnSerLybAgsLeuHisSerValArgThrSerSerGly 493
QY 1486 ACCGCCCAAGAGTCCCTTTATCTCCACATGCTGCAATCGAAGTGGAGCGCTGAGCAATTAC 1545
DB 494 ThrAlaGlnArgSerLeuLeuSerProLeuHisAgsSerGlySerValGluApsAenTy 513
QY 1546 CACACAGCTTCTTTTGGGCTTTCACACAGCGAGCAGCTACAGAAAGCTGTGCTGGC 1605
DB 514 HisTrpSerPheLeuPheGlyLeuSerThrSerGlnHisLeuThrLybSerAlaGly 533
QY 1606 CTGGGCTTAAAGGCTGCGACACTCGGATATCTTGAGCCCGCCAGACCTTACCCCTTCCCTG 1665
DB 534 LeuGlyLeuLybGlyTrpHisSerAps1leLeuAlaProGlnThrSerThrProSerLeu 553
QY 1666 ACCACAGCTGATTTTGGCCACAGAGTCTTCAACACTTACTGCTCAGCAGCATTCAC 1725
DB 554 ThrSerSerTrpTybPheAlaThrGluSerSerHisPheTySerAlaSerAla1leTy 573
QY 1726 GAGGCGAGTCCAGTTACTCTGCTTCAAGCTGACAGCGACGCTCCACTGGCGAGACCA 1785
DB 574 GlyLybSerAlaSerTySerAlaTySerCybSerGlnLeuProThrCybGlyApsGln 593
QY 1786 GTCTATTCTGTCGCGAGCGCGCAGAAAGCCAAAGTGAACAGACTGACTCGCGCGAGCTGG 1845
DB 594 ValTySerValArgArgArgGlnLybProSerApsArgAlaApsSerArgAgsTrp 613
QY 1846 CATGAGAGAGCCCTTTGAAAAGCATTTAAACGCAAGAGCTGCCAAATGGAATTTGCA 1905
DB 614 HisGluGluSerProPheGluLybGlnPheLybAgsArgSerCybGlnMetGluPheGly 633
QY 1906 GAGAGCATTAATGTCAGAAACAGTCAACGGGAAGCTGGGGAAGTGGGCGAGTCACTCT 1965
DB 634 GluSer1leMetSerGluApsArgSerArgGluGluLeuGlyValGlySerGlnSer 653
QY 1966 ACCTTTTCGCGCAGCATGGAATCATTTGAGTCTCC 2001
DB 654 SerPheSerGlySerMetGlu1leGluValSer 665

RESULT 5
US-10-072-012-681
/ Sequence 681, Application US/10072012
/ Publication No. US20040033493A1
/ GENERAL INFORMATION:
/ APPLICANT: Tchernev, Velizar
/ APPLICANT: Spytek, Kimberly
/ APPLICANT: Zelnusen, Bryan
/ APPLICANT: Patcurajan, Meera
/ APPLICANT: Shinkets, Richard
/ APPLICANT: Li, Li
/ APPLICANT: Gangolli, Esba
/ APPLICANT: Padigaru, Muralidhara
/ APPLICANT: Anderson, David W.
/ APPLICANT: Raselli, Luca
/ APPLICANT: Miller, Charles E.
/ APPLICANT: Gerlach, Valerie
/ APPLICANT: Taupier Jr, Raymond J.
/ APPLICANT: Gusev, Vladimir Y.
/ APPLICANT: Colman, Steven D.
/ APPLICANT: Wolenc, Adam R.
/ APPLICANT: Pena, Carol E. A
/ APPLICANT: Furtak, Katarzyna
/ APPLICANT: Grose, William M.
/ APPLICANT: Alsbrook II, John P.
/ APPLICANT: Lepley, Denise M.
/ APPLICANT: Rieger, Daniel K.
/ APPLICANT: Burgess, Catherine E.
/ TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
/ FILE REFERENCE: 21402-258
/ CURRENT APPLICATION NUMBER: US/10/072,012
/ PRIOR FILING DATE: 2002-01-31
/ PRIOR APPLICATION NUMBER: 60/265,102
/ PRIOR FILING DATE: 2001-01-30
```

```
/ PRIOR APPLICATION NUMBER: 60/265,514
/ PRIOR FILING DATE: 2001-01-31
/ PRIOR APPLICATION NUMBER: 60/265,517
/ PRIOR FILING DATE: 2001-01-31
/ PRIOR APPLICATION NUMBER: 60/265,412
/ PRIOR FILING DATE: 2001-01-31
/ PRIOR APPLICATION NUMBER: 60/265,395
/ PRIOR FILING DATE: 2001-01-31
/ PRIOR APPLICATION NUMBER: 60/266,406
/ PRIOR FILING DATE: 2001-02-02
/ PRIOR APPLICATION NUMBER: 60/266,767
/ PRIOR FILING DATE: 2001-02-05
/ PRIOR APPLICATION NUMBER: 60/267,057
/ PRIOR FILING DATE: 2001-02-07
/ PRIOR APPLICATION NUMBER: 60/266,975
/ PRIOR FILING DATE: 2001-02-07
/ PRIOR APPLICATION NUMBER: 60/267,459
/ PRIOR FILING DATE: 2001-02-08
/ Remaining Prior Application data removed - See file wrapper or PAM.
/ NUMBER OF SEQ ID NOS: 1391
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 681
/ LENGTH: 665
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-072-012-681

Alignment Scores:
Pred. No.: 0 Length: 665
Score: 472.00 Matches: 472
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 70.76% Indels: 0
DB: 12 Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x US-10-072-012-681 (1-665)
QY 586 ACCCTGCAAGAGCTGACTTTATCCCGAGTCTCATTTCTCGTGCTGCAATGAC 645
DB 194 ThrCysProIyPProAspPheIleProGluSerHisPheLeuAArgValProValaIenAsp 213
QY 646 AGCTTTTGAGAAATTTTCCGCGTGTTGACAAATCAGTAGATTCTTGAAGAAAGA 705
DB 214 SerPheCysGluIySleuPProIleuPProIleuAspIySerValAspPheIleGluIySala 233
QY 706 AAAGCTCCATAGATGTGTTAGTCACTGTTAGCTGGATCTCCCGCTCCGCCACC 765
DB 234 LysAlaSerAsnGlyCysValIleuValHisCysIleuAlaGlyIleSerArgSerAlaThr 253
QY 766 ATCGCTATCGCTACATCATGAAGAGATGACATGCTTTAGATGAAGCTTACAGATT 825
DB 254 IleAlaIleAlaIyIleMetIySarIyMetAspMetSerIleuAspGluAlaIyArgPhe 273
QY 826 GTGAAAGAAAAAGACCTACTATATCTCAAACTTCTGAGGCTGCTGAGAC 885
DB 274 ValIySgIuIyArgPProThrIleSerProAsnPheAsnPheLeuGlyGlnIleuIenAsp 293
QY 886 TATGAGAGAAAGATTAAAGAACAGACTGAGATCAAGGCGCAAAAGCAAACTCAAGCTG 945
DB 294 TyrGluIySlySleIyAsnGlnThrGlyAlaSerGlyProIySerIySleuIySleu 313
QY 946 CTGCACTGAGAGAACCAATGAACTGCTGCTGCTGCTGAGAGGTTGACAGAAAAGC 1005
DB 314 LeuHisIleuGluIyPProAsnGluProValProAlaValSerGluGlyGlnIySleuSer 333
QY 1006 GAGAGCGCCCTCAGTCCACCTGTGCGCACTCTGTACTCTGAGAGCAGCAAGCAAGG 1065
DB 334 GluThrProIeuserProProCysAlaAspSerAlaIenSerGluAlaIaIAGlyGlnArg 353
QY 1066 CCGGCGATCCCGCGAGGCTGCCAGCGTCCAGCGTCCAGCGTCCAGCGTCCGCTGTTAAGAAC 1125
DB 354 ProValHisProIaSerValProSerValProSerValGlnProSerIleuIenGluAsp 373
```

```
QY 1126 AGCCCGCTGTACAGAGCGCTCAGTGGCTGACCTGTCCGACAGAGCGTGAAGACAGC 1185
DB 374 SerProIeuserValGlnAlaIleuSerIyIleuHisIleuSerAlaAspArgIleuGluAspSer 393
QY 1186 AATAAGCTCAAGCGTTCTCTCTGTGATATCAAAATCAGTTTCATATTACCCAGCAGT 1245
DB 394 AsnIySleuIyArgSerPheSerIleuAspIleIySerSerValSerIySerAlaSerMet 413
QY 1246 GCAGCATCTTAAGAGGCTTCTCCCTCATCAGAAAGTCTTTGGAATCTATCAAACTTCC 1305
DB 414 AlaAlaSerIleuHisGlyPheSerSerSerIySerGluAspAlaIleuGluIyIyIyIyProSer 433
QY 1306 ACTACTGTGATGGAGCAACAGATATGCAAGTCTCTCCCTGTTCAAGAACTATACGAG 1365
DB 434 ThrThrIleuAspGlyThrAsnIySleuIySgIenPheSerProValGlnIleuSerGlu 453
QY 1366 CAGACTCCGAAACCAAGTCTGTATAGAGAGAAAGCCAGCATCCCGCAAGCTGCAAGC 1425
DB 454 GlnThrProGluThrSerProAspIySgIuAlaSerIleProIyIySleuGlnThr 473
QY 1426 GCCAGGCTTCAGACAGCAGCAGAGGATTTGCTTGGTCCAGAACCAAGCAGCAGTGGC 1485
DB 474 AlaArgProSerAspSerGlnSerIySarIyGluHisSerValArgThrSerSerSerGly 493
QY 1486 ACCGCCAGAGGTCCTTTATCTCCACCTGATCGAAGATGAGGAGCGTGAAGCAATTAC 1545
DB 494 ThrAlaGlnArgSerIleuSerProIeuserProIeuserHisArgSerGlySerValGluAspAsnIy 513
QY 1546 CACACAGCTTCTTTTCCGCTTTTCCACAGCAGCAGCAGCAGCAGCTTCAAGTCTGCTGGC 1605
DB 514 HisThrSerPheIleuPheGlyIleuSerThrSerGlnGlnHisIleuThrIySerAlaIy 533
QY 1606 CTGGGCGCTTAAGGGCTGCACTCGGATCTTTGGCCCCCGGACCTCAACCTTCCCTG 1665
DB 534 LeuGlyIleuIySgIyTrpHisSerAspIleuAlaProGlnThrThrProSerIleu 553
QY 1666 ACCAGCAGCTGTATTTTGGCAGACAGTCTCACACTTCTACTGCTGCTCAGCAGCATCAC 1725
DB 554 ThrSerSerTrpIyPheAlaIleuThrGluSerSerHisPheIySerAlaIleIy 573
QY 1726 GAGGCGAGTGCAGTACTCTGCTTAACAGCTGACAGCAGCTGCCACTTGGCGAGACCA 1785
DB 574 GlyIySerAlaSerIyIySerAlaIySerCysSerGlnIleuProThrCysGlyAspGln 593
QY 1786 GTCTATTCTGTGCCAGCGGCGCAGAAAGCAAGTCAAGACTGACTGCGCGGAGACTGG 1845
DB 594 ValIySerValArgArgArgGlnIyProSerAspArgAlaAspSerArgArgSerTrp 613
QY 1846 CATGAGAGAGAGCCCTTTGAAGAAAGCAATTAAAGCAGAGCTGCCAAATGGAATTGGA 1905
DB 614 HisGluIleuSerProPheGluIyGlnPheIyArgArgSerCysGlnMetGluPheGly 633
QY 1906 GAGAGCATCATGTGAGAAACAGGTCAAGGAGAGAGCTGGGAGAAAGTGGCGACTCAGTCT 1965
DB 634 GluSerIleuMetSerGluAsnArgSerArgGluIleuGluIyValGlySerGlnSer 653
QY 1966 AGCTTTCCGCGCAGACATGAAATCATTTAGAGTCTCC 2001
DB 654 SerPheSerGlySerMetGluIleIleGluValSer 665

RESULT 6
US-10-168-506-14
/ Sequence 14, Application US/10168506
/ Publication No. US20040053229A1
/ GENERAL INFORMATION:
/ APPLICANT: PLOWMAN, GREGORY D.
/ APPLICANT: MARTINEZ, RICARDO
/ APPLICANT: WHYTE, DAVID
/ APPLICANT: MANNING, GERARD
/ APPLICANT: SUDARSANAM, SUCHA
/ APPLICANT: HILL, RON
/ APPLICANT: FLANAGAN, PETER
/ TITLE OF INVENTION: MAMMALIAN PROTEIN PHOSPHATASES
```



```
/ FILE REFERENCE: 038602/1351
/ CURRENT APPLICATION NUMBER: US/10/168,506
/ CURRENT FILING DATE: 2002-06-21
/ PRIOR APPLICATION NUMBER: PCT/US00/34736
/ PRIOR FILING DATE: 2000-12-21
/ NUMBER OF SEQ ID NOS: 76
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 14
/ LENGTH: 665
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-168-506-14

Alignment Scores:
Pred. No.: 0 Length: 665
Score: 472.00 Matches: 472
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 70.76% Indels: 0
DB: 12 Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x US-10-168-506-14 (1-665)

QY 586 ACCTGTCCAAAGCCTGACTTATCCCGAGTCTCATTTCCGCGTGTGCTGAATGAC 645
DB 194 ThrcyProlyserProaspheHleProGluSerHlePheLeuAlaGValProValaAsnAsp 213
QY 646 AGCTTTTGGACAAATTTTGGCGTGTGACAAATTCAGTAGATTTCATTGAGAAAGCA 705
DB 214 SerPheCyGluLysIleLeuProTyrLeuAspLysSerValaAspPheIleGluLysVala 233
QY 706 AAAGCTCCAAATGATGTGTTTCAAGTGCATGTTTGGCTGGGATCTCCGCTCCGCGCAC 765
DB 234 LysAlaSerAsnLysValLeuValHisCybLeuAlaGlyLysSerArgSerAlaThr 253
QY 766 ATGCTATCGCTTACATCATGAGAGATGACATGCTTTAGATGAAGCTTACAGATT 825
DB 254 IleAlaIleAlaIleTyrIleMetLysArgMetAspMetSerLeuAspGluLalaTyrArgPhe 273
QY 826 GTGAAAGAAAAGACCTTATATCTCCAAATTCATTTTGGGCGCACTCTGAGAC 885
DB 274 ValLysGluLysArgProThrIleSerProAsnPheAsnPheLeuLysIleLeuLeuAsp 293
QY 886 TATGAGAAGAAGATTAGAACAAGATGAGATGAGGCGCAAGAGCAAACTCAGACGTG 945
DB 294 TyrGluLysLysIleLysAsnGlnThrGlyAlaSerGlyProLysSerLysLeuLysLeu 313
QY 946 CTGACCTGGAGAGCAATGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1005
DB 314 LeuHisLeuGluLysProAsnGluProValProAlaValSerGluLysIleGluLysSer 333
QY 1006 GAGACGCCCCCTGAGTCCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1065
DB 334 GluThrProLysSerProProCybAlaAspSerAlaThrSerGluLalaIleGluLysArg 353
QY 1066 CCGGTGCAATCCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1125
DB 354 ProValHisProAlaSerValProSerValProSerValGluProSerLeuLeuGluLys 373
QY 1126 AGCGGCTGGTACAGCGGCTGAGTGGGCTGACCTGCTGCGGAGACAGGCTGGAGAGACAG 1185
DB 374 SerProLeuValGluLalaLeuSerGlyLeuHisLeuSerAlaAspArgLeuGluLysSer 393
QY 1186 AATAAGCTCAAGCGTTCTTCTCTGTGATATCAATCAAGTTTCATATTTCAAGCAGACGT 1245
DB 394 AsnLysLeuLysArgSerPheSerLeuAspIleLysSerValSerTyrSerAlaSerMet 413
QY 1246 GCGAGCATCTTAATAGCTTCTCTCATGAGAGAGCTTTTGAATAACTACAACTTTC 1305
DB 414 AlaAlaSerLeuHisIleLysPheSerSerSerGluAspAlaLeuGluLysTyrTyrSerPro 433
QY 1306 ACTACTCTGGATGGAGCAACAAAGCTATGCAAGTTTCCCGCTGTTCAAGAACTACGAGAG 1365
```

```
DB 434 ThrThrLeuAspGlyThrAsnLysLeuCybGlnPheSerProValGlnLysLeuSerGlu 453
QY 1366 CAGACTCCCGGAAGACAGTCTGTATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1425
DB 454 GlnThrProGluThrSerProAspLysGluLysAlaSerIleProLysLysLeuGlnThr 473
QY 1426 GCCAGGCTTTCAGAGAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1485
DB 474 AlaArgProSerAspSerGlnSerLysArgLeuHisLeuSerValArgThrSerSerSerGly 493
QY 1486 ACCGCGCAGAGGTCCTTTTATCTCCACTGATCGAAGTGGAGCGGTGAGAGCAATTAC 1545
DB 494 ThrAlaGlnArgSerLeuLeuSerProLeuHisArgSerGlySerValGluLysAsnTyr 513
QY 1546 CACACAGCTTCTTTTCCGCTTTCACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1605
DB 514 HisThrSerPheLeuPheIleLeuSerThrSerGlnGlnHisLeuThrLysSerAlaGly 533
QY 1606 CTGGGCTTAAAGGCTGGGCTGCGATCGATATCTTGGAGCGGCGGAGAGAGAGAGAGAGAG 1665
DB 534 LeuGlyLeuLysGlyTyrPheIleSerAspIleLeuAlaProGlnThrSerThrProSerLeu 553
QY 1666 ACCAGCAGCTGGTATTTTGGCAAGATCTTCACTTCTGCTGCTGCTGCTGCTGCTGCTGCT 1725
DB 554 ThrSerSerTyrTyrPheAlaThrGluSerSerHisPheTyrSerAlaSerAlaIleTyr 573
QY 1726 GGAGGAGTGGCAGATTACTCTGCTTACAGCTGACAGAGAGAGAGAGAGAGAGAGAGAGAG 1785
DB 574 GlyGlySerAlaSerTyrSerAlaTyrSerCybSerGlnLeuProThrGlyLysArgGln 593
QY 1786 GTCTATTCTGTGCGAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1845
DB 594 ValTyrSerValArgArgGlnLysProSerAspArgAlaAspSerArgArgSerTyr 613
QY 1846 CATGAAGAGAGCCCTTTGAAAAGCATTTAAACGACAGAGAGAGAGAGAGAGAGAGAGAGAG 1905
DB 614 HisGluLysSerProPheGluLysGlnPheLysArgArgSerCybGlnMetGluPheGly 633
QY 1906 GAGAGCATCATTCAGAGAACAGTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1965
DB 634 GluSerIleMetSerGluAsnArgSerArgGluGluLeuGlyLysValGlySerGlnSer 653
QY 1966 AGCTTTTGGGCGAGCATGGAATCATTTGAGGCTCC 2001
DB 654 SerPheSerGlySerMetGluIleIleGluValSer 665

RESULT 7
US-10-343-357-7
/ Sequence 7, Application US/10343357
/ Publication No. US20040058341A1
/ GENERAL INFORMATION:
/ APPLICANT: INCYTE GENOMICS, INC.; TANG, Y. Tom
/ APPLICANT: ELIOTT, Vicki S.; RAMKOMAR, Jayalaxmi
/ APPLICANT: YAO, Monique G.; BURFORD, Neil
/ APPLICANT: WANG, Yumei E.; STEWART, Elizabeth A.
/ APPLICANT: GANDHI, Ameena R.; ARVIZU, Chandra S.
/ APPLICANT: LEE, Ernestine A.; HAFALITA, April J.A.
/ APPLICANT: LU, Dying Aina M.; TRIBOULEY, Catherine M.
/ APPLICANT: GRIFFIN, Jennifer A.; BAUGHN, Mariah R.
/ APPLICANT: YUE, Henry; WARREN, Bridget A.
/ APPLICANT: NGUYEN, Danielle B.; CHAWLA, Narinder K.
/ APPLICANT: KEARNEY, Liam
/ TITLE OF INVENTION: PROTEIN PHOSPHATASES
/ FILE REFERENCE: PI-0173 PCT
/ CURRENT APPLICATION NUMBER: US/10/343,357
/ CURRENT FILING DATE: 2003-01-28
/ PRIOR APPLICATION NUMBER: PCT/US01/23716
/ PRIOR FILING DATE: 2001-07-26
/ PRIOR APPLICATION NUMBER: US 60/221,679
/ PRIOR FILING DATE: 2000-07-28
/ PRIOR APPLICATION NUMBER: US 60/223,272
/ PRIOR FILING DATE: 2000-08-03
/ PRIOR APPLICATION NUMBER: US 60/224,309
```

; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: US 60/226,728
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: US 60/229,254
; PRIOR FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: US 60/231,366
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PERL Program
; SEQ ID NO: 7
; LENGTH: 665
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040058341A1 7480570CD1
US-10-343-357-7

Alignment Scores:

Pred. No.: 0 Length: 665
Score: 472.00 Matches: 472
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 70.76% Indels: 0
DB: 12 Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x US-10-343-357-7 (1-665)

QY 586 ACCGTGCGCAAGCTGATCTTATCCCGAGTCTCATTTCCGTGCTGCTGTAATGAC 645
DB 194 TttCyProLyserProAspRheIleProGlnSerHisPheLeuValArgValProValIaenAsp 213
QY 646 AGCTTTTGAGAAATTTTGGCCGTGTGGAACAATCAGTATGATTTCTTGAAGAAAGA 705
DB 214 SerPheCysGlnLysIleLeuProTrpLeuAspLysSerValAspPheIleGlnLysAla 233
QY 706 AAAGCTCCAAATGATGTGTTCTAGTGACTGTTTAGCTGGATCTCCCGCTCCGACAC 765
DB 234 LysAlaSerAsnGlyCysValLeuValHisCysLeuAlaGlyIleSerArgSerAlaThr 253
QY 766 ATGCGTATGCGCTACATCATGAAAGAGATGACATGCTTTAGATGAAGCTTACAGATT 825
DB 254 IleAlaIleAlaIleArgIleMetLysArgMetAspMetSerLeuAspGlnAlaIleArgPhe 273
QY 826 GTGAAGAAAGAAAGAACTGATATCTGCAAACTTCAATTTCTGGGCCAACTCGTGGAC 885
DB 274 ValLysGlnLysArgProThrIleSerProAsnPheAsnPheLeuGlyGlnLeuLeuAsp 293
QY 886 TATGAGAGAAGATTAAAGAACAGACTGAGATCAGAGCCAAAGAACTCAAGCTG 945
DB 294 TyrGlnLysLysIleLysAsnGlnThrGlyAlaSerGlyProLysSerLysLeuLysLeu 313
QY 946 CTGACCTGAGAGAGCCAAATGAATGACTGTCCTGCTGCTGCTCAGAGGGTGAAGAAAGC 1005
DB 314 LeuHisIleGlnLysProAsnGlnProValProAlaValSerGlnLysGlnLysSer 333
QY 1006 GAGAGCCCTCAGTCCACCCCTGTGCGGACTGCTGACTGACGAGACACAGCAAAAG 1065
DB 334 GlnThrProLeuSerProProCysAlaAspSerAlaThrSerGlnAlaIleGlyGlnArg 353
QY 1066 CCCGTCATCCCGCAGCGTGCACCGTGCAGCTGCGCAGGTGAGCCGTGCTTGAAGAC 1125
DB 354 ProValHisProAlaSerValProSerValProSerValGlnProSerLeuLeuGlnAsp 373
QY 1126 AGCCGCTGATACAGCGCTCAGTGGCTGACCTGTCCGACAGACAGGCTGGAAGACAC 1185
DB 374 SerProLeuValGlnAlaLeuSerGlyLeuHisIleLeuSerAlaAspArgLeuGlnAspSer 393
QY 1186 AATTAAGCTAAGGCTCTTCTCTGATATCAATCAATGTTTCAATTCAGCGACAGC 1245
DB 394 AsnLysLeuLysArgSerPheSerIleuAspIleLysSerValSerTyrSerAlaSerMet 413
QY 1246 GCAGATCTCTTACATGAGCTTCTCTCATGACAGAGATGCTTTGGAATATCAAACTTCC 1305

DB 414 AlaAlaSerLeuHisGlyPheSerSerSerGlnAspAlaLeuGlnLysTyrTyrLysProSer 433
QY 1306 ACTACTCTGATGGAGCAACAAGCTATGCCAGTTCTCCCTGTTCAAGAACTACGGAG 1365
DB 434 ThrThrLeuAspGlyThrAsnLysLeuCysGlnPheSerProValGlnGlnLeuSerGln 453
QY 1366 CAGACTCCCGAAACAGTCTGATGATGAGAGGAAGCAACATCCCAAGAAAGTGCAGACC 1425
DB 454 GlnThrProGlnThrSerProAspLysGlnGlnAlaSerIleProLysLysLeuGlnThr 473
QY 1426 GCCAGCCTTCAGACAGCCAGACGAAGCATTCATTGCTGACAGACAGACAGTGGC 1485
DB 474 AlaArgProSerAspSerGlnSerLysArgLeuHisSerValArgThrSerSerGly 493
QY 1486 ACCGCCCAAGAGTCCCTTTTATCTTCACATGCAATGCAATGGAGAGGTGAGAGCAATTAC 1545
DB 494 ThrAlaGlnArgSerLeuLeuSerProLeuHisIleAspSerGlySerValGlnAspAsnTyr 513
QY 1546 CACACACAGCTTCCTTTGCGCCTTTCACACGACGACGACACCTCAGCAAGTCTGTGAC 1605
DB 514 HisTrpSerPheLeuPheGlyLeuSerThrSerGlnGlnHisLeuThrLysSerAlaGly 533
QY 1606 CTGGGCTTAAAGGCTGCGACCTCGGATATCTTGCGCCCCAGACCTTACCCCTTCCCTG 1665
DB 534 LeuGlyLeuLysGlyTyrPheIleSerAspIleLeuAlaProGlnThrSerThrProSerLeu 553
QY 1666 ACCAGCAGCTGATATTTTGGCAGACAGATCCACACTTCTACTGCTGCGCTGATAC 1725
DB 554 ThrSerSerTyrPyrPheAlaThrGlnSerSerHisPheTyrSerAlaSerAlaIleTyr 573
QY 1726 GGAGCAGATGCGAGTACTGCTGCTTACAGCTGACGACGCGCCACTTGCGGAGACCA 1785
DB 574 GlyGlySerAlaSerTyrSerAlaTyrSerCysSerGlnLeuProThrCysGlyLysAspGln 593
QY 1786 GTTATTTCTGCGCAGCGCGGAGAGCCAAAGTGAACAGACTGACTCGCGGCGAGCTGG 1845
DB 594 ValTyrSerValArgArgArgGlnLysProSerAspArgAlaAspSerArgArgSerTyrP 613
QY 1846 CATGAAGAGAGCCCTTTGAAAAGCATTTAAAGCAAGCTGCGCAATGAAATTTGGA 1905
DB 614 HisGlnLysSerProPheGlnLysGlnPheLysAspArgSerCysGlnMetGlnPheGly 633
QY 1906 GAGACATCATGTCCAGAAACAGTCAAGGAGAAAGTGGGAAAGTGGGAGTCACT 1965
DB 634 GlnSerIleMetSerGlnAsnArgSerArgGlnGlnLeuGlyLysValGlySerGlnSer 653
QY 1966 AGCTTTTGGGCGAGCATGAAATCATTGAGGTTCTCC 2001
DB 654 SerPheSerGlySerMetGlnIleIleGlnValSer 665

RESULT 8
US-10-094-749-2312
; Sequence 2312, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUKIO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOKYU
; APPLICANT: NAGAHARI, KENJI

; APPLICANT: MASUHO, YASUHIKO
 ; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
 ; FILE REFERENCE: 084335/0160
 ; CURRENT APPLICATION NUMBER: US/10/094,749
 ; CURRENT FILING DATE: 2002-03-12
 ; PRIOR APPLICATION NUMBER: 60/350,435
 ; PRIOR FILING DATE: 2002-01-24
 ; PRIOR APPLICATION NUMBER: JP 2001-328381
 ; PRIOR FILING DATE: 2001-09-14
 ; NUMBER OF SEQ ID NOS: 3381
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 2312
 ; LENGTH: 665
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-10-094-749-2312

Alignment Scores:

Pred. No.:	0	Length:	665
Score:	472.00	Matches:	472
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	70.76%	Indels:	0
DB:	15	Gaps:	0

US-10-029-345a-108_COPY_532_2532 (1-2001) x US-10-094-749-2312 (1-665)

QY 586 ACCTGTCCAAAGCTTGAATATCCCGAGCTTCATTCCTGCGTGTGTAATGAC 645
 DB 194 ThCyProlyserProaspheilleProgluSerHispheleuArgValProValaamap 213
 QY 646 AGCTTTTGTGAAGAAATTTTGGCGTGTGGAACAATTCAGTAATTTTCATGGAAGCA 705
 DB 214 SerPheCySeGluYserIleuProTyrLeuAspLysSerValAspPheIleGluYala 233
 QY 706 AAGCCCTCCCAATGATGTGTGTAGTCACTGTTAGCTGGAGTCCCGCTCCGACAC 765
 DB 234 LysAlSerSerHisCyValLeuValHisCySeuAlaGlyIleSerHisSerAlaThr 253
 QY 766 ATGCTATCGCTTACATCATGAAGAGATGACATGCTTTAGATGAAGCTTACAGATT 825
 DB 254 IleAlaIleAlaIleTyrIleMetLysArgMetAspMetSerLeuAspGluAlaTyrArgPhe 273
 QY 826 GTGAAGAAAGAAAGCTTATATATCTCCAACTTCAATTTTCTGGCCCACTCTGAC 885
 DB 274 ValLysGluYserArgProThrIleSerProAspPheAspPheLeuGluIleuLeuAsp 293
 QY 886 TATGAGAAGAAAGATTAAAGACAGACTGAGATCAGGCGCAAGCAAACTCAAGCTG 945
 DB 294 TyrGluYserLysIleYserAspGlnThrGlyAlaSerGlyProLysSerLysLeuYaleu 313
 QY 946 CTGACCTGGAGAAGCCAAATGAACCTGTCTCTGCTGTCTCAGAGGGTGGACAGAAAGC 1005
 DB 314 LeuHisLeuGluYserProAspGlnProValProAlaValSerGluGlyGlnLysSer 333
 QY 1006 GAGAAGCCCTCAGTCCACCTGTGCGACTTGTCTACTCAGAGGACAGAGCAAAAG 1065
 DB 334 GluThrProLeuSerProProCyAlaAspSerAlaThrSerGluAlaIlaGlyGlnArg 353
 QY 1066 CCGGTGATCCCGCCAGCGCGCCAGGCTCCAGCGGTGAGCGCTGCGTTAGAGGAC 1125
 DB 354 ProValHisProAlaSerValProSerValProSerValGlnProSerLeuLeuGluAsp 373
 QY 1126 AGCCCGTGTAGAGCGCTCAGTGGGCTGACCTGTCCGACAGAGCTGGAAGACAGC 1185
 DB 374 SerProLeuValGlnAlaLeuSerGlyLeuHisLeuSerLysAspArgLeuGluAspSer 393
 QY 1186 AATTAAGTCAAGCGTTCTTCTCTCTGTGATATCAATCAAGTTTCATATTCAGCCAGATG 1245
 DB 394 AsnLysLeuYserArgSerPheSerLeuAspIleLysSerValSerTyrSerAlaSerMet 413
 QY 1246 GCAGCATCTTACATGCTTCTCTCTCTCATCAGAGAGCTTTGGAATACTCAAACTTCC 1305

DB 414 AlaAlaSerLeuHisGlyPheSerSerSerGluAspAlaLeuGluTyrTyrLysProSer 433
 QY 1306 ACTACTCTGGATGGAGCAAGGATATGCTCCCTGCTCCGAGCAATATGCGAG 1365
 DB 434 ThrThrLeuAspGlyThrAsnLysLeuCyGlnPheSerProValGlnIleuSerGly 453
 QY 1366 CAGACTCCCGAAACAGTCTCTGATAGAGAGAGAGCAGCATCCCAAGAGTGCAGACC 1425
 DB 454 GlnThrProGluThrSerProAspLysGluGluAlaSerIleProLysLysLeuGlnThr 473
 QY 1426 GCCAGGCTTCAAGACAGCCAGCAAGCAGATTGCAATTCGTCAGAACCCAGCAGCTGCC 1485
 DB 474 AlaArgProSerAspSerGlnSerLysArgLeuHisSerValArgThrSerSerGly 493
 QY 1486 ACCGCCCAAGGCTCCCTTTATCTCCACATGCTCCAAAGTGGAGGCTGAGCAATTC 1545
 DB 494 ThrAlaGlnArgSerLeuSerProLeuHisArgSerGlySerValGluAspAsnTyr 513
 QY 1546 CACACAGCTTCTTTTGGCTTTCACACAGCCAGCAGCAGCAGCTCAGAACTGCTGGC 1605
 DB 514 HisThrSerPheLeuPheGlyLeuSerThrSerGlnGlnHisLeuThrLysSerAlaGly 533
 QY 1606 CTGGGCTTAAAGGCTGGCAGCTCGGATATCTTGCCCCCAGACCTTACCCCTTCCTG 1665
 DB 534 LeuGlyLeuLysGlyTyrHisSerAspIleLeuAlaProGlnThrSerThrProSerLeu 553
 QY 1666 ACCAGAGCTGTATTTTTCACAGAGTCTTACACTTACTCTGCTCCAGCATCTAC 1725
 DB 554 ThrSerSerTyrTyrPheAlaThrGluSerSerHisPheTyrSerLysSerAlaIleTyr 573
 QY 1726 GGAGGAGAGCCAGTACTCTGCTTACAGCTCAGCAGCTCCCACTTGGCGAGACAA 1785
 DB 574 GlyGlySerAlaSerTyrSerAlaTyrSerCySeGlnLeuProThrCySeGlyAspGln 593
 QY 1786 GTCTATTCTGTGCGAGCGCGCAGAGCCAAAGTACAGAGCTGACTCGCGGAGACTGG 1845
 DB 594 ValTyrSerValArgArgArgGlnLysProSerAspArgAlaAspSerArgArgSerTyr 613
 QY 1846 CATGAAGAGAGCCCTTTGAAGAGAGCTTAAACGCAAGAGCTGCCAAATGGAATTGGA 1905
 DB 614 HisGluGluSerProPheGluYserGlnPheLysArgArgSerCySeGlnMetGluPheGly 633
 QY 1906 GAGAGCATCATCAGAGAAAGGTCAGCGGAAGAGCTGGGGAAGTGGCAGTCACTCT 1965
 DB 634 GluSerIleMetSerGluAsnArgSerArgGluGluLeuGlyLysValIleYserGlnSer 653
 QY 1966 AGCTTTTCGGCAGCATGGAATCATTTGAGTCTCC 2001
 DB 654 SerPheSerGlySerMetGluIleIleGluValSer 665

RESULT 9
 US-10-377-072-26
 ; Sequence 26, Application US/10377072
 ; Publication No. US20040009501A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Millennium Pharmaceuticals Inc.
 ; APPLICANT: Curtis, Rory A.J.
 ; APPLICANT: Logan, Thomas Joseph
 ; APPLICANT: Glucksmann, Maria A.
 ; APPLICANT: Meyers, Rachel E.
 ; APPLICANT: Williamson, Mark J.
 ; APPLICANT: Rudolph-Owen, Laura A.
 ; APPLICANT: Chun, Miyoung
 ; APPLICANT: Tsai, Feng-Ying
 ; TITLE OF INVENTION: NOVEL 25869, 25934, 26335, 50365, 21117,
 ; TITLE OF INVENTION: 38692, 46508, 16816, 16839, 49937, 49931 AND 49933 MOLECULAR
 ; FILE REFERENCE: MP103-0180NMIM
 ; CURRENT APPLICATION NUMBER: US/10/377,072
 ; PRIOR FILING DATE: 2003-02-27
 ; PRIOR APPLICATION NUMBER: US 09/895,860
 ; PRIOR FILING DATE: 2001-06-29
 ; PRIOR APPLICATION NUMBER: US 60/215,370

```

; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: US 09/723,806
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: US 60/187,455
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: US 09/843,297
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: US 60/199,801
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: US 09/861,801
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: US 60/205,508
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: US 09/816,494
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/815,419
; PRIOR FILING DATE: 2001-03-22
; Remaining prior application data removed - See file wrapper or PAM.
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 26
; LENGTH: 665
; TYPE: PRT
; ORGANISM: Homo Sapiens
; US-10-377-072-26

Alignment Scores:
pred. No.: 0 Length: 665
Score: 472.00 Matches: 472
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 70.76% Indels: 0
DB: 15 Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x US-10-377-072-26 (1-665)
QY 586 ACCTGTCAGAAAGCTGCTTATCCCGAGTCTCATTTCTGCGTGGCTGTAATGAC 645
DB 194 ThrCysProLeuSerProAspPheIleProGlnSerHisPheLeuArgAlaProValAsnAsp 213
QY 646 AGCTTTTGTGAGAAATTTTGGCCGTGGTGACAAATCAAGTGAATTTCAATTGAGAAAGA 705
DB 214 SerPheCysGlnLysIleLeuProTLeuAspLysSerValAspPheIleGlnLysAla 233
QY 706 AAGCCTCCAGTGGATGTTCTTACGACCTGTTTGGTGGATCTCCCGCTCCGCGAC 765
DB 234 LysAlaSerAsnGlnCysValLeuValHisCysLeuAlaGlyIleSerIrgSerAlaTr 253
QY 766 ATGCTATGCGCTTACATCATGAAGAGATGACATGCTTTAGATGAAGCTTACAGATT 825
DB 254 IleAlaIleAlaIleTyrIleMetLysArgMetAspMetSerLeuAspGlnAlaTyrArgPhe 273
QY 826 GTGAAAGAAAAAGACCTTATATCTCAAACTTCAATTTTCTGGGCCAACTCCTGAGAC 885
DB 274 ValLysGlnLysArgProThrIleSerProAsnPheAsnPheLeuGlnLeuLeuAsp 293
QY 886 TATGGAAGAAGATTAGAACAAGATGAGATGACGGGCCAAAGGCAAACTCAAGCTG 945
DB 294 TyrGlnLysLysIleLysAsnGlnTrpGlyAlaSerGlyProLysSerLysLeuLysLeu 313
QY 946 CTGCACTTGAGAGAGCAATATGAACCTGTCCCTGCTGTCTGACAGGGGTGAGAAAGC 1005
DB 314 LeuHisLeuGlnLysProAsnGlnLysProValProAlaValSerGlnLysGlnLysSer 333
QY 1006 GAGAGCCCCCTGACCTGCTGTGGCGCACTGTCTACTCTCAGAGGACAGAGCAAG 1065
DB 334 GlnTrpProLeuSerProProCysAlaAspSerAlaThrSerGlnAlaIleGlnArg 353
QY 1066 CCGCGGCAATCCCGCGAGCGTGGCCAGCGTGGCCAGGTCGACCGCTGTTAGAGAC 1125
DB 354 ProValHisProAlaSerValProSerValProSerValGlnProSerLeuGlnLys 373
QY 1126 AGCCCGCTGTACAGGCGCTCAGTGGCTGACCTGTCCGACAGACGCTGAGAAAGAC 1185
```

```

DB 374 SerProLeuValGlnAlaLeuSerGlyLeuHisLeuSerAlaAspArgLeuGlnLysAsp 393
QY 1186 AATAAGCTCAAGCGTCTCTCTCTGAGATATCAAAATGATTTTCAATTATCCCGCGACATG 1245
DB 394 AsnLysLeuLysArgSerPheSerLeuAsnLysIleLysSerValSerTyrSerAlaSerMet 413
QY 1246 GCAGCATCTTACATGAGCTCTCCCTCATCAGAAAGCTTTGGAATATCTACAACTTCC 1305
DB 414 AlaAlaSerLeuHisGlyPheSerSerSerGlnAlaLeuGlnLysTyrIleLysProSer 433
QY 1306 ACTACTGTGATGGAGCAACAAGCTATGCCAGTTCTCCCTGTTACAGAACTATGGAG 1365
DB 434 ThrThrLeuAspGlyThrAsnLysLeuCysGlnPheSerProValGlnGlnLysSerGln 453
QY 1366 CAGATCCCGAAACCAAGTCTGATTAAGAGAGAGACACCATCCCAAAAGCTGAGACC 1425
DB 454 GlnTrpProGlnLysSerProAspLysGlnGlnAlaSerIleProLysLysLeuGlnTrp 473
QY 1426 GCCAGGCTTACAGACGACGAGCAAGGATTTGATTCGTCAGAACCCAGACGATGGAC 1485
DB 474 AlaArgProSerAspSerGlnSerIleLysArgLeuHisSerValArgTrpSerSerGly 493
QY 1486 ACCGCGCAAGAGTCCCTTTTATCTTCACATGCATGCAAGTGGAGCGGTGAGACAATTAC 1545
DB 494 ThrAlaGlnArgSerLeuLeuSerProLeuHisArgSerGlySerValGlnLysAsnTrp 513
QY 1546 CACACACACTTCTTTTGGGCTTTCCACAGCCGACGACGACCTCAGCAAGTCTGCG 1605
DB 514 HisThrSerPheLeuPheGlyLeuSerHisGlnGlnHisLeuThrLysSerAlaGly 533
QY 1606 CTGGGCTTAAAGGGGTGGACCTCGGATATCTTGGCCCCCAGACCTTACCCCTTCCCTG 1665
DB 534 LeuGlyLeuLysGlyTyrHisSerAspIleLeuAlaProGlnTrpSerThrProSerLeu 553
QY 1666 ACCAGACCTGTATTTTGGACACAGAGTCTTCACTTCTATCTGCTCAGACCATTTAC 1725
DB 554 ThrSerSerTrpLysPheAlaThrGlnLysSerHisPheTyrSerAlaSerAlaIleTyr 573
QY 1726 GAGGCACTGCCAGTACTCTGCTCAGCTGACGCTGCGACCCAGCTGCCACTTGGGAGACCA 1785
DB 574 GlyLysSerAlaSerTyrSerAlaTyrSerCysSerGlnLysProThrCysGlyAspGln 593
QY 1786 GTCTATTCTGTGGGAGGCGGAGAAAGCCAAAGTGAACAGCTGACTCGCGGAGAGCTGG 1845
DB 594 ValTyrSerValArgArgArgGlnLysProSerAspArgAlaAspSerArgArgSerTrp 613
QY 1846 CATGAAGAGACCCCTTTGAAAAGCATTTTAAAGCAGAACTGCAATGGAATTTTGA 1905
DB 614 HisGlnLysSerProPheGlnLysGlnPheLysArgArgSerCysGlnMetGlnPheGly 633
QY 1906 GAGAGCATCATGTGAGAAACAGTCAACGCGGAAAGCTGGGAAAGTGGGAGTCACTCT 1965
DB 634 GlnSerIleMetSerGlnAsnArgSerArgGlnLysLeuGlyLysValGlySerGlnSer 653
QY 1966 AGCTTTGCGGAGCATGAATCATTTGAGGCTCC 2001
DB 654 SerPheSerGlySerMetGlnIleIleGlnValSer 665

RESULT 10
US-10-257-026-2
; Sequence 2, Application us/10257026
; Publication No. US20040086659A1
; GENERAL INFORMATION:
; APPLICANT: Merck Patent GmbH
; TITLE OF INVENTION: New dual specificity phosphatase
; FILE REFERENCE: DUSP10XDM5
; CURRENT APPLICATION NUMBER: US/10/257,026
; CURRENT FILING DATE: 2003-11-07
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 665
```

```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-029-345a-108_538_2532.01go.rapb
Alignment Scores:
Pred. No.: 0 Length: 665
Score: 472.00 Matches: 472
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 70.76% Indels: 0
DB: 16 Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x US-10-257-026-2 (1-665)

QY 586 ACCTGTCGCAAGAGCTGATTTATCCCGAGTCTCATTTCTGCGTGGCTGCTGAATGAC 645
DB 194 ThnCysProlyserProbaPheHleProGluSerHlePheLeuArgValProValAaPhe 213
QY 646 AGCTTTGTGAGAAATTTTGGCGGTGAGCAAAATCAGTATGATTTTCATTTGAGAAAGCA 705
DB 214 SerPheCysGluValIleLeuProTyrLeuAspLysSerValAaPheHleGluValAa 233
QY 706 AAAGCTCCCAATGATGTGTTCAGTGTGACCTGTTAGCTGGAATCTCCGCGCCGAC 765
DB 234 LysAlaSerAenglyCysValLeuValHleCysLeuHlaGlyIleSerArgSerAlaThr 253
QY 766 ATGGCTATGCGCTTCATCATGAAAGAGATGACATGCTTTAGTGAAGCTTACAGATTT 825
DB 254 IleAlaIleAlaTyrIleMetLysArgMetAspMetSerLeuAspGluAlaTyrArgPhe 273
QY 826 GTGAAAGAAAAGACCTATATATCTCAAACTTCAATTTTGGGCGCAACTCTGAGAC 885
DB 274 ValLysGluValArgProHlePheSerProAsnPheAsnPheLeuGlyIleLeuLeuAsp 293
QY 886 TATGAGAAAGATTAAGAACAGACTGAGACATCAAGGCGCAAGAACTCAAGCTG 945
DB 294 TyrGluValLysIleLysValAenglyIleThrGlyAlaSerClyProLysSerLysLeuVal 313
QY 946 CTGCACTGAGAAAGCAAAATGAATGATCTGCTGCTGCTCAAGAGGTGAGCAAAAGC 1005
DB 314 LeuHleValLeuGluValLysProAenglyProValProAlaValSerGluGlyIleLysSer 333
QY 1006 GAGAGCGCCCTCAGTCAACCTGCGGCGACTGCTGCTCAAGAGGAGAGGAGCAAAAG 1065
DB 334 GluThrProLeuSerProProCysValAaPheSerAlaThrSerGluAlaAlaGlyIleAsp 353
QY 1066 CCGGTGATCCCGCAGCGCTGCCAGCGTCCCGAGCTGAGCGCTGTTAGAGAC 1125
DB 354 ProValHleProAlaSerValProSerValProSerValGlnProSerLeuLeuGluAsp 373
QY 1126 AGCCCGCTGTACAGCGGCTCAGTGGGCTGACCTGTCCGAGACAGAGCTGAAAGACAGC 1185
DB 374 SerProLeuValGlnAlaLeuSerGlyLeuHleValSerAlaAaPheArgLeuGluAspSer 393
QY 1186 AATAAGCTCAAGGGTCTCTCTCTGAGATATCAATCAATTCATTTTACAGCAGAG 1245
DB 394 AsnLysLeuLysArgSerPheSerLeuAspLleLysSerValSerLysSerAlaSerMet 413
QY 1246 GCAGCATCTTACATGAGCTTCTCTCATCAGAGATGCTTTGAAATTAATAACCTTCC 1305
DB 414 AlaAlaSerLeuHleGlyPheSerSerSerSerGluAspAlaLeuGluValTyrLysProSer 433
QY 1306 ACTACTCTGAGAGGAGCAAAAGACTATGCAAGTCTCCCTGTTCAAGAACTATCGAG 1365
DB 434 ThrThrLeuAspGlyThrAsnLysLeuCysGlnPheSerProValGlnGluLeuSerGly 453
QY 1366 CAGACTCCGAAACCGACTCTGATAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1425
DB 454 GlnThrProGluThrSerProAspLysGlnGluAlaSerLleProLysLysValLeuGlnThr 473
QY 1426 GCCAGGCTTCAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1485
DB 474 AlaArgProSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 493

```

```

QY 1486 ACCGCCAGAGGACCTTTTATCTCCAGTGCATGCAAGTGGAGCGGAGAGCAAAATAC 1545
DB 494 ThrAlaGlnArgSerLeuLeuSerProLeuHleArgSerGlySerValGluAspAsnTyr 513
QY 1546 CACACAGCTTCCTTTTTCGCGCTTTCCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1605
DB 514 HisThrSerPheLeuHleGlyLeuSerThrSerGlnGlnHleValThrLysSerAlaGly 533
QY 1606 CTGGGCTTAAAGGCTGAGCATCTGATATCTTGGCCCGCCAGACCTTACCTTCCCTG 1665
DB 534 LeuGlyLeuLysGlyTyrPheSerAspLleLeuAlaProGlnThrSerThrProSerLeu 553
QY 1666 ACCAGAGCTGGATTTTTCACAGAGCTCTCCACATTTACTCTGCTCCAGCATCTAC 1725
DB 554 ThrSerSerTyrPheHleAlaThrGluSerSerHlePheLysSerAlaSerAlaIleTyr 573
QY 1726 GGAGCAGTGCAGTACTCTGCTCAGCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1785
DB 574 GlyGlySerAlaSerLysSerAlaTyrSerCysSerGlnLeuProThrCysGlyAspGln 593
QY 1786 GTCTATCTGTGGCGAGCGCGGAGAGAGCAAGTGCAGAGCTGACTGCGCGGAGCTGG 1845
DB 594 ValTyrSerValArgArgArgGlnLysProSerAspArgAlaAspSerArgArgSerTyr 613
QY 1846 CATGAGAGAGCGCCCTTTGAAAAGAGCTTTAAACGAGAGAGCTGCCAAATGGAATTGGA 1905
DB 614 HisGluGluSerProHleGluLysGlnPheLysValArgArgSerGlyIleMetGluPheGly 633
QY 1906 GAGAGCATCATGTACAGAGAAAGAGTCAAGGAGAGAGCTGGAGAAAGTGGGAGTCACT 1965
DB 634 GluSerIleMetSerGluAsnArgSerArgGluGluLeuGlyLysValGlySerGlnSer 653
QY 1966 AGCTTTGCGGAGCATGGAATCATTTGAGTCTCC 2001
DB 654 SerPheSerGlySerMetGluIleIleGluValSer 665

RESULT 11
US-10-648-593-240
; Sequence 240, Application US/10648593
; Publication No. US20040106132A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION OF GENES FOR PREDICTING ACTIVITY OF COMPOUNDS THAT
; TITLE OF INVENTION: INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE KINASES AND/OR
; FILE REFERENCE: D0273 NP
; CURRENT APPLICATION NUMBER: US/10/648,593
; PRIOR FILING DATE: 2003-08-26
; PRIOR FILING DATE: 2002-08-27
; NUMBER OF SEQ ID NOS: 557
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 240
; LENGTH: 665
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-648-593-240
Alignment Scores:
Pred. No.: 0 Length: 665
Score: 472.00 Matches: 472
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 70.76% Indels: 0
DB: 16 Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x US-10-648-593-240 (1-665)

QY 586 ACCTGTCGCAAGAGCTGATTTATCCCGAGTCTCATTTCTGCGTGGCTGCTGAATGAC 645
DB 194 ThnCysProlyserProbaPheHleProGluSerHlePheLeuArgValProValAaPhe 213

```

OY	646	AGCTTTTGAGAAAATTTTTCGCGGTGGAGCAAACTCAGTAAATTTTCATTGAGAAAACA	705
Db	214	SerPheCysGluIysIleLeuProItrPheuSpIysSerValaIaPheIleGluIysAla	233
OY	706	AAAGCCTTCAAATGAGATGTGTTCTAGTGCACTGTTTAGCTGGAGATCTCCGCTCCGCCAC	765
Db	234	LysIaIaSerAengIyCysValIeuValIhIscYsIeuIaIagIyIleSerAysSerAlaIthr	253
OY	766	ATCCCTATCCGCTCACTCATCATGAAGAAGATGCACTAGCTTTAGATGAAAGCTTAACAATTT	825
Db	254	IleAlaIleAlaIyTlIeMeLysArgMetCaspMetSerLeuAspGluAlaIyAArgPhe	273
OY	826	GTGAAGAAAAAAGACCTCACTATATCTCCAAACTTCATATTTTCGGGCCAATCCTTGAC	885
Db	274	ValIysGluIysArgProThrIleSerProAsnPheAsnPheLeuGluIleuLeuAsp	293
OY	886	TATGAGAAAGAAATTAAGAACCCAGACTGGAGCATCAGGGCCAAAGAGCAAACTCAAGCTG	945
Db	294	TyrGluIysIysIleIysAsnGluThrIyAlaSerGlyProIySerIySleuIyIleu	313
OY	946	CTGCAACCTCGAGGAAGCCAAATGAACCTGTCCCTGTGTCTCAGAGGGGCGACAGAAAAGC	1005
Db	314	LeuIhIleuGluIysProAsnGluProValProAlaValSerGluGlyGluIlySer	333
OY	1006	GAGAGCGCCCTCAATGCCACCTGTGCGCACTTGTCTACTCTCAGAGCGACAGACAAGG	1065
Db	334	GluThrProIeuserProProCysAlaAspSerAlaThrSerGluAlaIaIagIyGlnArg	353
OY	1066	CCCGGATCACTCCGCGAGGTCGCCAGAGTCGCCAGCGCGAGCCGTGCTGTAGAGAC	1125
Db	354	ProValhIspProIaSerValProSerValProSerValGlnProSerLeuIeueGluAsp	373
OY	1126	AGCCCGCTGTATACAGCGCCTCAGTGGGCTGCACCTGTCCGAGACAGGCTTGAAGACAGC	1185
Db	374	SerProIeuValIglAlaIeuserClyIeuIhIleuserAlaIaAspArgIeugluAspSer	393
OY	1186	AATAAGCTCAAGCGTCTCTTCTCTGTGATATCAAAATCAGTTTCATATTCAGCCAGATG	1245
Db	394	AsnIySleuIysArgSerPheSerIeueuAspIleIySberValIserIySerAlaIaSerIet	413
OY	1246	GCAGATCCTTAACATGCTTCTCCCTCCATCAAGAAGTCTTTGAAATCTACAACCTTCC	1305
Db	414	AlaIaIaSerIeuehIaGlyPheSerIeSerIeuIuAspAlaIeueGluIyTylIyIeProSer	433
OY	1306	ACTACTCTGATGGGACCAACAAGCTATGCCAGTTCCTCCCTGTTCAGAACTATCGGAG	1365
Db	434	ThrThrIeueAspGlyThrAsnIySleuIySgIinPheSerProValIglIngluIeuserCly	453
OY	1366	CAGACTCCCGGAAACCAAGCTGTATTAAGAGGAAGCACAATCCCAAGAACTGACAGAC	1425
Db	454	GlnhItrProGluIhIhSerProAspIySgluGluAlaSerIleProIySlyIeueGluIhItr	473
OY	1426	GCCAGGCTTCAGACAGGCCAGAGCAAGCAGATTGCGATGGGTCAAGAACCGAGCAGATGGC	1485
Db	474	AlaArgProSerAAspSerGlnSerIySArgIeuhIaSerValIArgThrSerSerIeCly	493
OY	1486	ACCGGCCAGAGATCCCTTTATATTCACATGCATGCAATGGGAGCGTGGAGAGCAATTAC	1545
Db	494	ThrAlaGlnArgSerIeueuserProIeuhIaArgSerGlySerValGluAspAsnIy	513
OY	1546	CACACCAACTTCCTTTTCGAGCTTTCACACAGCAGCAGCAGCACTCAACAAGTCTGTGGC	1605
Db	514	HIsThrSerPheIeuePheGlyIeuserThrSerGlnGlnhIleuThrIySerAlaIeIy	533
OY	1606	CTGGGCTTAAAGGCTGGGCACTCGGATATCTTTGGCCCCCAGACTCTAACCCCTTCCCTG	1665
Db	534	LeuGlyIeueIySgIyTlRPIaSerAAspIleIeueAlaProGluIhIhSerThrProSerIeu	553
OY	1666	ACAGACACTGGTATTTTGGCACAAGCTCCACACTTCTACTGCTCAGCATCTAC	1725
Db	554	ThrIserSerIyPlyPheAlaIhIhIuIeuserhIhIspheIyIySerAlaIaIeIy	573
OY	1726	GGAGGACAGTCCCACTTACTCTGCTTACAGCTGGACCGACTGCCCACTTGCAGAGACCA	1785

QY	Db	QY	Db	QY	Db	QY	Db	QY	Db
1786	574	1786	574	1786	574	1786	574	1786	574
GTCAATTCGTGGCCAGGCGGCAGAAAGCAAGTGCACAGAGCTGACTCGCGGGGAGCTGG	GlyGlySerAlaSerIleYrSerAlaTyrSerCysSerGlnLeuProThrCysGlyAspGln	GTCAATTCGTGGCCAGGCGGCAGAAAGCAAGTGCACAGAGCTGACTCGCGGGGAGCTGG	GlyGlySerAlaSerIleYrSerAlaTyrSerCysSerGlnLeuProThrCysGlyAspGln	GTCAATTCGTGGCCAGGCGGCAGAAAGCAAGTGCACAGAGCTGACTCGCGGGGAGCTGG	GlyGlySerAlaSerIleYrSerAlaTyrSerCysSerGlnLeuProThrCysGlyAspGln	GTCAATTCGTGGCCAGGCGGCAGAAAGCAAGTGCACAGAGCTGACTCGCGGGGAGCTGG	GlyGlySerAlaSerIleYrSerAlaTyrSerCysSerGlnLeuProThrCysGlyAspGln	GTCAATTCGTGGCCAGGCGGCAGAAAGCAAGTGCACAGAGCTGACTCGCGGGGAGCTGG	GlyGlySerAlaSerIleYrSerAlaTyrSerCysSerGlnLeuProThrCysGlyAspGln
594	594	594	594	594	594	594	594	594	594
ValIyrSerValAlaGhArgGlnIlyrSerThrSphArgAlaAspSerArgIyrSerTrp		ValIyrSerValAlaGhArgGlnIlyrSerThrSphArgAlaAspSerArgIyrSerTrp		ValIyrSerValAlaGhArgGlnIlyrSerThrSphArgAlaAspSerArgIyrSerTrp		ValIyrSerValAlaGhArgGlnIlyrSerThrSphArgAlaAspSerArgIyrSerTrp		ValIyrSerValAlaGhArgGlnIlyrSerThrSphArgAlaAspSerArgIyrSerTrp	
1846	1846	1846	1846	1846	1846	1846	1846	1846	1846
CATGAAGAGAGCCCTTTGAAAGCAGTTTAAAGCCAGAACTGCGCAATGGAATTGGGA		CATGAAGAGAGCCCTTTGAAAGCAGTTTAAAGCCAGAACTGCGCAATGGAATTGGGA		CATGAAGAGAGCCCTTTGAAAGCAGTTTAAAGCCAGAACTGCGCAATGGAATTGGGA		CATGAAGAGAGCCCTTTGAAAGCAGTTTAAAGCCAGAACTGCGCAATGGAATTGGGA		CATGAAGAGAGCCCTTTGAAAGCAGTTTAAAGCCAGAACTGCGCAATGGAATTGGGA	
614	614	614	614	614	614	614	614	614	614
HisGlnGlnSerProPheGlnIlyGlnPheIlyrSphArgSerCysGlnMetGlnIlyGly		HisGlnGlnSerProPheGlnIlyGlnPheIlyrSphArgSerCysGlnMetGlnIlyGly		HisGlnGlnSerProPheGlnIlyGlnPheIlyrSphArgSerCysGlnMetGlnIlyGly		HisGlnGlnSerProPheGlnIlyGlnPheIlyrSphArgSerCysGlnMetGlnIlyGly		HisGlnGlnSerProPheGlnIlyGlnPheIlyrSphArgSerCysGlnMetGlnIlyGly	
1906	1906	1906	1906	1906	1906	1906	1906	1906	1906
GAGAGCATCATGTGAGAGAACAGSTCAAGGAGAGAGCTGGGAAAGTGGGAGTCAGTCT		GAGAGCATCATGTGAGAGAACAGSTCAAGGAGAGAGCTGGGAAAGTGGGAGTCAGTCT		GAGAGCATCATGTGAGAGAACAGSTCAAGGAGAGAGCTGGGAAAGTGGGAGTCAGTCT		GAGAGCATCATGTGAGAGAACAGSTCAAGGAGAGAGCTGGGAAAGTGGGAGTCAGTCT		GAGAGCATCATGTGAGAGAACAGSTCAAGGAGAGAGCTGGGAAAGTGGGAGTCAGTCT	
634	634	634	634	634	634	634	634	634	634
GlnSerIleMetSerGlnAlaAsnArgSerArgGlnGlnIleuGlyIlyValGlySerGlnSer		GlnSerIleMetSerGlnAlaAsnArgSerArgGlnGlnIleuGlyIlyValGlySerGlnSer		GlnSerIleMetSerGlnAlaAsnArgSerArgGlnGlnIleuGlyIlyValGlySerGlnSer		GlnSerIleMetSerGlnAlaAsnArgSerArgGlnGlnIleuGlyIlyValGlySerGlnSer		GlnSerIleMetSerGlnAlaAsnArgSerArgGlnGlnIleuGlyIlyValGlySerGlnSer	
1966	1966	1966	1966	1966	1966	1966	1966	1966	1966
AGCTTTTCGGCAGACATGGAATCATATGAGGCTCC	2001	AGCTTTTCGGCAGACATGGAATCATATGAGGCTCC	2001	AGCTTTTCGGCAGACATGGAATCATATGAGGCTCC	2001	AGCTTTTCGGCAGACATGGAATCATATGAGGCTCC	2001	AGCTTTTCGGCAGACATGGAATCATATGAGGCTCC	2001
654	654	654	654	654	654	654	654	654	654
SerPheSerGlySerMetGlnIleIleGlnValSer	665	SerPheSerGlySerMetGlnIleIleGlnValSer	665	SerPheSerGlySerMetGlnIleIleGlnValSer	665	SerPheSerGlySerMetGlnIleIleGlnValSer	665	SerPheSerGlySerMetGlnIleIleGlnValSer	665

Db 294 TyrGluYulYbYlleylYbAenGlnThrGlyAlaSerGlyProLYSerLYbLeuYLeu 313
QY 946 CTGCACCTGGAGAGCAAAATGAACTGTCCCTGTCTCTCAGAGGCTGGACGAAGC 1005
Db 314 LeuHlYleuYlYbProAenGluProValProAlaValSerGluGlyGlnLYSer 333
QY 1006 GAGACGCCCTTCAGTCACCCCTGTGCGACTGTCTGCTCACTCAGAGGAGAGCAAG 1065
Db 334 GluThrProLeuSerProProGlybAlaSerSerAlaThrSerGluAlaAlaGlyGlnG 353
QY 1066 CCCGTGCATCCCGCAGCGCTGCCAGCGTCCAGCGCTGCAGCGCTGTAGAGAC 1125
Db 354 ProValHlYbProAlaSerValProSerValProSerValGlnProSerLeuGluAaP 373
QY 1126 AGCCCGCTGTGTACAGGGCTTCAGTGGCTCAGCTGTCCCGCAGACAGCTGGAAAGAC 1185
Db 374 SerProLeuValGlnAlaLeuSerGlyLeuHlYbLeuSerAlaAspArgLeuGluAaP 393
QY 1186 AATAGCTCAAGCGTTCCTCTCTGTGATATCAATCATGTTTCATATTCAGCGAGATG 1245
Db 394 AsnLYbLeuYbArgSerPheSerLeuAaPrlLYbSerValSerLYbSerAlaSerMet 413
QY 1246 GCAGCATCTTACATGGCTTCTCTCATCAGAAAGATGCTTGGAAATACAAACCTTCC 1305
Db 414 AlaAlaSerLeuHlYbGlyPheSerSerSerGlyAaPrlAlaLeuGluYlYrYlYbProSer 433
QY 1306 ACTACTGTGATGGAGCAACAAGCTATGCGCAGTTCCTCCCTGTTCAGAAATATCGGAG 1365
Db 434 ThrThrLeuAaPrlYbThrAaLYbLeuCYbGlyPheSerProValGlnGluLeuSerGlu 453
QY 1366 CAGACTCCGAAACAGACTCTGATTAAGAGAAAGCCAGCTCCCGCAAGAGCTGACAGAC 1425
Db 454 GlnThrProLuhTrSerProAaPrlYbGluGluAlaSerLYbLeuYlYbLeuGlnThr 473
QY 1426 GCCAGGCTTTCAGACAGCAAGCAAGCAAGATGCTGATTCGCTCAGAAACAGCAGCTG 1485
Db 474 AlaArgProSerAaPrlSerGlnSerLYbArgLeuHlYbSerValArgThrSerSerGly 493
QY 1486 ACCGCCCAAGAGTCCCTTTATATCTCCATGATCGAAGTGGAGCGTGGAGCAATTAC 1545
Db 494 ThrAlaAlaArgSerLeuLeuSerProLeuHlYbArgSerLYbSerValGluAaPrlY 513
QY 1546 CACACGAGCTTCCTTTTGGGCTTTCCACAGCAGCAGAGCAAGCTCCAGAGCTGCGC 1605
Db 514 HlYbThrSerPheLeuPheGlyLeuSerThrSerGlnHlYbLeuThrLYbSerAlaGly 533
QY 1606 CTGGGCTTAAAGGCTGGCACTCGATATCTTGGCCCCCAGCACTTACCCCTTCCCTG 1665
Db 534 LeuGlyLeuLYbGlyTrpHlYbSerAaPrlLeuAlaProLinhTrSerThrProSerLeu 553
QY 1666 ACCAGCAGCTGTGATTTTGGCAAGAGTCTCAGACTTACTGCTGCTCAGCCATCTAC 1725
Db 554 ThrSerSerTrpYlYbPheAlaTrpGluSerSerHlYbPheYlYbSerAlaSerAlaLeY 573
QY 1726 GAGAGCGAGGCGCACTTCTGCTCAGCTCAGCTGCAAGCTGCGCCACTTGGAGAGCCAA 1785
Db 574 GlyLYbSerAlaSerLYbSerLYbSerAlaYlYbSerLYbLeuProThrCYbGlyAaPrl 593
QY 1786 GTCATATCTGTGGCAGAGCGCAGAAAGCAAGTGAAGAGCTGACGCGCGAGCTGG 1845
Db 594 ValTYbSerValArgArgGlnLYbProSerAaPrlArgAlaPheSerArgAaPrlSerTrp 613
QY 1846 CATGAAGAGAGCCCTTTGAAAAGAGTAAAGCAGAAAGCTGCCAAATGAAATTTGGA 1905
Db 614 HlYbGluLuhTrSerProPheGlyLYbGlnPheLYbArgArgSerCYbGlnMetGluPheGly 633
QY 1906 GAGAGCATCATGTCAGAGAGCAGGTCACGGAGAGAGCTGGGGAGAAAGTGGCACTGCTC 1965
Db 634 GluSerLYbMetSerGluAaPrlArgSerArgGluGluLeuLYbValGlySerGlnSer 653
QY 1966 AGCTTTTGGGAGCATGAATCATTTAGGCTTCC 2001
Db 654 SerPheSerGlySerMetGluLlYlYbGlnValSer 665

RESULT 13
US-10-072-012-679
Sequence 679, Application US/10072012
Publication No. US20040033493A1
GENERAL INFORMATION:
APPLICANT: Tchernyev, Velizar
APPLICANT: Spyrek, Kimberly
APPLICANT: Zethusen, Bryan
APPLICANT: PatkuraJan, Weera
APPLICANT: Shimkete, Richard
APPLICANT: Li, Li
APPLICANT: Gangolli, Beha
APPLICANT: Padigaru, Muralidhara
APPLICANT: Anderson, David W.
APPLICANT: Rastelli, Luca
APPLICANT: Miller, Charles E.
APPLICANT: Gerlach, Valerie
APPLICANT: Taupier Jr, Raymond J.
APPLICANT: Gusev, Vladimír Y.
APPLICANT: Colman, Steven D.
APPLICANT: Wolenc, Adam R.
APPLICANT: Pena, Carol E. A
APPLICANT: Furtak, Kacarzyna
APPLICANT: Grosse, William M.
APPLICANT: Alsobrook II, John P.
APPLICANT: Lepley, Denise M.
APPLICANT: Rieger, Daniel K.
APPLICANT: Burgess, Catherine E.
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-258
CURRENT APPLICATION NUMBER: US/10/072,012
CURRENT FILING DATE: 2002-01-31
PRIOR APPLICATION NUMBER: 60/265,102
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: 60/265,514
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,517
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,412
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,395
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/266,406
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 60/266,767
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: 60/267,057
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/266,975
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/267,459
PRIOR FILING DATE: 2001-02-08
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 1391
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 679
LENGTH: 690
TYPE: PRN
ORGANISM: Homo sapiens
US-10-072-012-679
Alignment Scores:
Pred. No.: 0
Score: 472.00
Percent Similarity: 100.00%
Beet Local Similarity: 100.00%
Query Match: 70.76%
DB: 12
Length: 690
Matches: 472
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0
US-10-029-345A-108_COPY_532_2532 (1-2001) x US-10-072-012-679 (1-690)
QY 586 ACCTGTCCAAAGCCTGACTTATCCCGAGTCTCATTTCTGCGTGTGCTGGAATGAC 645

Db 219 ThCysProLysProAspPheIleProGlnSerHisPheLeuArgValProValIAsnAsp 238
Qy 646 AGCTTTTGAGAAATTTTGGCGTGGTGGACAAATCACTAATTTTCATTGGAAGCA 705
Db 239 SerPheCysGluLysIleLeuProTyrPLeuAspLysSerValAspPheIleGluLysAla 258
Qy 706 AAAGCCTCAATGAGATGTGTTCTAGTGCACCTGTTAGCTGGAGTCTCCGCTCCGACAC 765
Db 259 LysAlaSerAsnGlnCysValIleuValHisCysLeuAlaGlyIleSerArgSerAlaThr 278
Qy 766 ATGGCTATGGCTTCATCATGAAGAAGATGAACATGCTTTTAAATGAAGCTTACAGATT 825
Db 279 ILeAlaIleAlaTyrIleMetLysArgMetAspMetSerLeuAspGluAlaTyrArgPhe 298
Qy 826 GTGAAGAAGAAAGACCTATATCTCCAACTTGAATTTTGGGCGCAATCCCGAC 885
Db 299 ValLysGluLysArgProThrIleSerProAsnPheAsnPheLeuGlnLeuLeuAsp 318
Qy 886 TATGAGAAGAATTAAGAACAGACTGAGCATCAAGGCGCAAGAGCAAACTCAAGCTG 945
Db 319 TyrGluLysLysIleLysAsnGlnThrGlnLysSerGlyProLysSerLysLeuLysLeu 338
Qy 946 CTGCACCTGGAGAAAGCCAAATGAACCTGCTCTGCTGCTCAGAGGCTGAGCAAGAAC 1005
Db 339 LeuHisLeuGluLysProAsnGlnProValProAlaValSerGluGlyGlnLysSer 358
Qy 1006 GAGAGCGCCCTGAGTCAACCTGTGCGCACTGTCAACCTCAGAGGAGAGAGCAAG 1065
Db 359 GlnThrProLeuSerProProCysAlaAspSerAlaThrSerGlnAlaGlyGlnArg 378
Qy 1066 CCGGTGCATCCCGCAGCGTCCAGCGTCCAGCGGTGAGCGCTGCTTGAAGAGAC 1125
Db 379 ProValHisProAlaSerValProSerValProSerValGlnProSerLeuGlnAsp 398
Qy 1126 AGCCCGCTGGTAAAGCGCTCACTGAGGCTGCACCTGTCCGACAGACAGCTGAAGACAC 1185
Db 399 SerProLeuValGlnAlaLeuSerGlyLeuHisIleLeuSerAlaAspArgLeuGlnAspSer 418
Qy 1186 AATAAGCTCAAGGCTCTCTCTCTGAGATATCAATTCAGTTTCAATTCAGCCGACAG 1245
Db 419 AsnLysLeuLysArgSerPheSerLeuAspIleLysSerValSerTyrSerAlaSerMet 438
Qy 1246 GCAGCATCTTACATGAGCTTCTCTCATGAGAAAGATGCTTGGAACTATCAAACTTCC 1305
Db 439 AlaAlaSerLeuHisGlyPheSerSerSerGlnAspAlaLeuGlnLysTyrTyrLysProSer 458
Qy 1306 ACTACTCTGATGGAGCAACAAAGCTATGCGCACTTCTCCCTGTTGAGAACTATCGAG 1365
Db 459 ThrThrLeuAspGlyThrAsnLysLeuCysGlnPheSerProValGlnLysLeuSerGln 478
Qy 1366 CAGACTCCCGAAACAGCTCTGATTAAGAGAAAGCAGAGATCCCGCAAGAGCTGACAG 1425
Db 479 GlnThrProGlnThrSerProAspLysGlnGlnAlaSerIleProLysLysLeuGlnThr 498
Qy 1426 GCGAGGCTTTCAGACAGCAGCAAGCAGATTCGATTCGATGAGAACAGCAGCAGTGC 1485
Db 499 AlaArgProSerAspSerGlnSerLysArgLeuHisSerValArgHisSerSerSerGly 518
Qy 1486 ACCGCCCAAGGTCCTTTTATCTCCATGCAATCGAAGTGGAGCGGTGAGACAATTAC 1545
Db 519 ThrAlaGlnArgSerLeuLeuSerProLeuHisArgSerGlySerValGlnAspAsnTyr 538
Qy 1546 CACACAGGTCCTTTTGGGCTTTTCCAGACCGACAGACACTTCAGAAAGCTGTGTCGC 1605
Db 539 HisThrSerPheLeuPheGlyLeuSerThrSerGlnGlnHisLeuThrLysSerAlaGly 558
Qy 1606 CTGAGCTTAAAGGCTGAGCTCGATATCTTGGCCCGCAGACCTTACCCCTTCCCG 1665
Db 559 LeuGlyLeuLysGlyTyrHisSerAspIleLeuAlaProGlnThrSerThrProSerLeu 578
Qy 1666 ACCAGCAGCTGATTTTGGCAAGAGTCTTCAACTTACTGCTCAGCCACTTAC 1725

Db 579 ThrSerSerTyrPheAlaThrGlnSerSerHisPheTyrSerAlaSerAlaIleTyr 598
Qy 1726 GGAGGAGAGCCAGATTACTCTGCTCCATACAGCTGCACCGCCACCTGGCGGAGACCA 1785
Db 599 GlyGlySerAlaSerTyrSerAlaTyrSerCysSerGlnLeuProThrCysGlyAspGln 618
Qy 1786 GTCTATTCTGTCGCGAGCGCGCAGAGCCAGTGAACAGTCAAGCTACTCCGCGCGAGCTGG 1845
Db 619 ValTyrSerValArgArgArgGlnLysProSerAspArgAlaAspSerArgArgSerTyr 638
Qy 1846 CATGAAGAGCCCTTTGAAAAGCATTTAAACGCAAGCTGCCAAATGGAATTTGCA 1905
Db 639 HisGlnLysSerProPheGluLysGlnPheLysArgArgSerCysGlnMetGluPheGly 658
Qy 1906 GAGACATCATGTCAGAAACAGGTCACGAGGAAAGCTGGGGAAGCTGGCAGTACGCT 1965
Db 659 GluSerIleMetSerGlnAsnArgSerArgGlnGluLeuGlyLysValGlySerGlnSer 678
Qy 1966 AGCTTTTGGGCGAGCATGGAATCATTTGAGGCTTCC 2001
Db 679 SerPheSerGlySerMetGluIleIleGlnValSer 690

RESULT 14
US-10-072-012-703
; Sequence 703, Application US/10072012
; Publication No. US2004003493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zernusen, Bryan
; APPLICANT: Patuturajan, Meera
; APPLICANT: Shinkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangolli, Esha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Colman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grosse, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine B.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459

```
/ PRIOR FILING DATE: 2001-02-08
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 1391
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 703
/ LENGTH: 690
/ TYPE: PR
/ ORGANISM: Homo sapiens
US-10-072-012-703

Alignment Scores:
Pred. No.: 0          Length: 690
Score: 472.00        Matches: 472
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 70.76%          Indels: 0
DB: 12                  Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-10-072-012-703 (1-690)

QY 586 ACCGTGTCMAAGCTGACTTATCCCGAGTCTCATTTCTGCGTGTGCTGTAATGAC 645
DB 219 ThCybProlybProAbpRheIIeProGluSerHISpHeuAlaGValProValaAbp 238
QY 646 AGCTTTGTGAGAAATTTTGGCGTGGTGGACAATCAAGTAGATTTCATTGAGAAAGA 705
DB 239 SerPhCySeGluYsAlleuProTrpLeuAbpLysSerValaAbpHeIIeGluYsAla 258
QY 706 AAGCCTCCAAATGATGTGTTTCTAGTGCACCTGTTAGCTGGAGTCTCCGCTCCGACCC 765
DB 259 LysAlaSerAbnGluYsValleuValHISpHeuAlaGValLysSerAlaHeIIe 278
QY 766 ATGCTATGCGCTTACATCATGAAGAGATGGACATGTTTATGATGAAGTTTCAATTT 825
DB 279 HeIIeAlaHeIIeAlaYsHeIIeYsArgMetAbpMetSerLeuAbpGluAlaYsArgPhe 298
QY 826 GTGAAGAAAAAGACTACTATATCTCCAAACTTCATTTTGGGGCCAACTCTGGAC 885
DB 299 ValLysGluYsArgProTrpHeIIeSerProAbpRheAbpHeuGluYsGlnLeuAbp 318
QY 886 TATGAGAAAGATTAAGAACACAGACTGAGCATCAGGCGCAAGAACTCAAGCTG 945
DB 319 TyrGluYsLeuYsHeIIeYsAbnGlnThnGluYsAlaSerGluYsProLysSerLysLeuYsLeu 338
QY 946 CTGCACCTGGAGAAAGCAATGAACCTGTCTCTGCTGTCTCAAGGCTGGACGAAAGC 1005
DB 339 LeuHISpHeuGluYsAbpRheAbnGlnProValaAbpSerGluYsGlnLysSer 358
QY 1006 GAGAGCGCCCTGAGTCCAGCTGTGCGGACTCTGTAACCTCAGAGGAGGAGCAAAAG 1065
DB 359 GluThrProLeuSerProProCybAlaAbpSerAlaThSerGluAlaAlaGluYsGlnArg 378
QY 1066 CCCGTCATCCCGCAGCAGCTGCGCAGAGCTGCCAGGCTGGAGCGCTGCTTATGAGAGC 1125
DB 379 ProValHISpHeuAlaSerValaProSerValaProSerValaGlnProSerLeuLeuGluAbp 398
QY 1126 AGCCCGCTGTGACAGCGCTCAGTGGCTCAGCTGTCCGACAGACAGCTGGAAGCAGC 1185
DB 399 SerProLeuValaGlnAlaLeuSerGluYsLeuHISpHeuSerAlaAbpArgLeuGluAbpSer 418
QY 1186 AATTAAGCTCAAGGTTCTCTCTCTGAGATATCAATCAAGTTTCATATTAGCAGACAG 1245
DB 419 AsnLysLeuYsArgSerPheSerLeuAbpHeIIeYsSerValaSerYsSerAlaSerMet 438
QY 1246 GCAGCATCTTACATGGCTTCTCTCATCAGAGATGCTTTGGAATATCAAACTTCC 1305
DB 439 AlaAlaSerLeuHISpHeuSerSerSerSerGluAbpAlaLeuGluYsTrpLysProSer 458
QY 1306 ACTACTCTGGATGGAGCAACAAGCTATGCCAGTTCTCCCTGTTCAAGAACTATGGAG 1365
DB 459 ThrThrLeuAbpGluYsThrAsnLysLeuCybGlnPheSerProValaGlnGluYsLeuSerGlu 478
QY 1366 CAAACTCCGCAAAACAAGTCTGATTAAGAGAGAAAGCAAGCATCCCAAGAAAGCTGCAAGC 1425
```

```
DB 479 GluThrProGluThrSerProAbpLysGluGluAlaSerLeuAbpProLysLysLeuGlnThr 498
QY 1426 GCCAGGCTTTCAGACAGCCGAGCAAGCAAGCATTTGCTATCGGTGAGAAACCAAGCACTGGC 1485
DB 499 AlaArgProSerAbpSerGlnSerLysArgLeuHISpHeuSerValaArgThSerSerGly 518
QY 1486 ACCGCGCAGAGGTCCTTTATCTCCATGCACTGCAAGTGGAGGCTGGAGCAATTAAC 1545
DB 519 ThrAlaGlnArgSerLeuLeuSerProLeuHISpHeuSerGlySerValaGluAbpAntyr 538
QY 1546 CACACCAAGCTTCTTTTGGCTTTTCCAGCAGCAGCAGCAGCAGCAGTCAAGTGTGTCGC 1605
DB 539 HisThrSerPheLeuHeuGluYsLeuSerThrSerGlnGlnHISpHeuThrLysSerAlaGly 558
QY 1606 CTGGGCTTAAAGGCTGGGACCTGCAATCTTGAGCCGCCAGACCTTCACTCCCTTCCCG 1665
DB 559 LeuGluYsLeuYsGluYsThrPheSerAbpHeIIeAlaProGlnThrSerThrProSerLeu 578
QY 1666 ACCAGAGCTGGATTTTGGCAAGAGTCTGCACTTCTAGCTGCTCAGGCATCTAC 1725
DB 579 ThrSerSerTrpYsPheAlaThrGluSerSerHISpHeuYsSerAlaSerAlaIleYs 598
QY 1726 GAGGACATGTCAGATTACTCTGCTTACAGCTGACAGCAGCAGCTGCCACTTGGCAGACAA 1785
DB 599 GlyGlySerAlaSerYsSerAlaYsSerCybSerGlnLeuProThrCybGluYsArgGln 618
QY 1786 GTCTATTTCTGGCGCAGCGCGCAAGAACAGTGAACAGTGAAGTGAAGTGAAGTGAAGTGA 1845
DB 619 ValYsSerValaArgArgArgGlnLysProSerAbpArgAlaAbpSerArgArgSerTrp 638
QY 1846 CATGAAGAGAGCCCTTTTAAAGCAGTTTAAACGAGAAAGCTGCCAAATGGAATTGGA 1905
DB 639 HisGluGlnSerProHeuGluYsGlnPheLysArgArgSerCybGlnMetGluPheGly 658
QY 1906 GAGAGCATCATGTCAGAGAAACAGTTCACGCGAAGAGCTGGGAAAGTGGGAGTCACTG 1965
DB 659 GluSerHeIIeMetSerGluAbnArgSerArgGluGluLeuGluYsValaGlySerGlnSer 678
QY 1966 AGCTTTGCGGACATGGAATCATTAAGAGTCTCC 2001
DB 679 SerPheSerGlySerMetGluHeIIeGluValSer 690

RESULT 15
US-10-425-114-54204
/ Sequence 54204, Application US/10425114
/ Publication No. US20040034888A1
/ GENERAL INFORMATION:
/ APPLICANT: Liu, Jindong
/ APPLICANT: Zhou, Yihua
/ APPLICANT: Kovalic, David K.
/ APPLICANT: Screen, Steven E
/ APPLICANT: Tabaska, Jack E
/ APPLICANT: Cao, Yongwei
/ TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
/ FILE REFERENCE: 38-21(53313)B
/ CURRENT FILING DATE: US/10/425,114
/ PRIOR FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 73128
/ SEQ ID NO 54204
/ LENGTH: 690
/ TYPE: PR
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: Clone ID: LIB4119-028-H6_F11.pcp
US-10-425-114-54204

Alignment Scores:
Pred. No.: 0          Length: 690
Score: 472.00        Matches: 472
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
```

Query Match: 70.76% Indels: 0
DB: 12 Gaps: 0
US-10-029-345a-108_COPY_532_2532 (1-2001) x US-10-425-114-54204 (1-690)

QY 586 ACTGTCGAAGGCTGACCTTATCCCGAGTCTCATTTCTGCGTGTGCTGTGAATGAC 645
DB ThCysProIysProAspPheIleProGluSerHisPheLeuArgValProValAsnAsp 238
QY 646 AGCTTTGTGAAGAAATTTTGCCTGCTGTGACAAATCAGTAAGATTTCATTGAGAAAGCA 705
DB SerPheCysGluIysIleLeuProTyrPLeuAspIysSerValAspPheIleGluIysAla 258
QY 706 AAAGCCTCAATGGATGTGTTCTAGTGCATGCTTTAGCTGGAGCTCCGCGCTCCGACAC 765
DB LysAlaSerHisnGlyCysValLeuValHisCysLeuAlaIleIleSerArgSerAlaThr 278
QY 766 ATGCTATGCGCTCATCATGAAGAGATGAGACATGCTTTAGATGAAGCTTACAGATT 825
DB ILeAlaIleAlaTyrIleMetLysArgMetAspMetSerLeuAspGluAlaTyrArgPhe 298
QY 826 GTGAAGAAAGAAAGAACTTATATCTCCAACTTCAATTTTCTGGGCCAACTCTGAC 885
DB ValLysGluIysArgProIleThrIleSerProAsnPheAsnPheLeuGluIleLeuAsp 318
QY 886 TATGAGAAAGATTAAGAACAGACTGAGCATCAGGCGCAAGAGCAAACTCAAGCTG 945
DB TyrGluIysLysIleLysAsnGlnThrGlyAlaSerGlyProIysSerIysLeuIysLeu 338
QY 946 CTGACACTGGAAGAACCAATGAACCTGTCTCTGCTGTCTCAGAGGCTGACAGAAAGC 1005
DB LeuHisLeuGluIysArgProAsnGluProValProAlaValSerGluIysGluIysSer 358
QY 1006 GAGAGCGCCCTGACGACCGCTGTGCGAGCTGTGCTACCTCAGAGGACAGAGCAAAAG 1065
DB GlnThrProLeuSerProProCysAlaAspSerAlaThrSerIleAlaIleGlyIysArg 378
QY 1066 CCCGTCATCCCGACGCTGCCAGCGTCCAGCGTCCGAGCGCTGCTGTTAGAGAGAC 1125
DB ProValHisProAlaSerValProSerValProSerValGlnProSerLeuLeuIysAsp 398
QY 1126 AGCCCGCTGTACAGCGCTCAGTGGGCTGACCTTCCGACAGACAGGCTGGAAGACAC 1185
DB SerProLeuValGlnAlaLeuSerGlyLeuHisIleLeuSerAlaAspArgLeuGluAspSer 418
QY 1186 AATAAGCTCAAGGCTTCTCTCTCTGATATCAATTCATTAATTCACCCGACAG 1245
DB AsnLysLeuLysArgSerPheSerIleuAspIleLysSerValSerTyrSerIleAsnMet 438
QY 1246 GAGACATCTTACATGAGCTTCTCTCATGACAGAGATGCTTTGGAATACTACAACCTTCC 1305
DB AlaAlaSerLeuHisGlyPheSerSerSerGluAspAlaLeuGluIleTyrTyrIysProSer 458
QY 1306 ACTACTCTGTGATGGGACCAACAAGCTATAGCCAGTTTCCCTGTTGAGAACTATGAGAG 1365
DB ThrThrLeuAspGlyThrAsnLysLeuCysGlnPheSerProValGlnIleuLeuSerGlu 478
QY 1366 CAGACTCCCGAAACAGTCTGTATAGAGAGAGAGCAGAGATCCCAAGAAAGCTGAGAGCC 1425
DB GlnThrProGluThrSerProAspLysGluGluAlaSerIleProLysLysLeuGlnThr 498
QY 1426 GCGAGGCTTCAAGACAGCAGAGCAAGCATTCATTCGCTGAGAACAGCAGCAGTGGAC 1485
DB AlaArgProSerAspSerGlnSerIysArgLeuHisIleSerValArgThrSerSerGly 518
QY 1486 ACCGCCAAGAGTCCCTTTATCTCACTGCATCGAAGTGGAGAGGCTGAGAGCAATTAC 1545
DB ThrAlaGlnArgSerIleuLeuSerProLeuHisAspSerGlySerValGluAspAsnLys 538
QY 1546 CACACAGGTTCTTTTGAGCTTTCACACAGCAGCAGACACTTCAGAAAGCTGCTGAGC 1605
DB HisThrSerPheLeuPheGlyLeuSerThrSerGlnGlnHisIleuThrLysSerAlaGly 558

QY 1606 CTGGGCGCTTAAGGCGTGGACATCGGATATCTTGGCCCCCAGACCTCTACCCCTTCCCTG 1665
DB LeuGlyLeuLysGlyTyrPheIleSerAspIleLeuAlaProGlnThrSerThrProSerLeu 578
QY 1666 ACCAGCAGCTGTGATTTTTCACAGAGTCCCTCAGACTTCTACTGCTGACCTCAGCATTCAC 1725
DB ThrSerSerTyrPheAlaThrGluSerHisPheTyrSerHisIleAlaIleTyr 598
QY 1726 GAGAGCAGTGGCAATTAATCTTGCTTACAGCTGACCCAGCTGCCCATTTGCCGAGACCA 1785
DB GlyGlySerAlaSerTyrSerAlaTyrSerCysSerGlnLeuProThrCysGlyAspGln 618
QY 1786 GTCTATTCTGTGCGAGGCGGAGAGCCAGTGAACAGCTGCATCCGCGCGGAGCTGG 1845
DB ValTyrSerValArgArgArgGlnLysProSerAspArgAlaAspSerArgArgSerTyr 638
QY 1846 CATGAAGAGAGCCCTTTGAAAGAGATTAAACCAAGCTGCCAATGGAATTTTGA 1905
DB HisGluGluSerProPheGluLysGlnPheLysArgArgSerCysGlnMetGluPheGly 658
QY 1906 GAGACATCATGTCAGAGAACAGTCAAGGAGAGCTGGGGAAAGTGGGCAATCATGCT 1965
DB GlnSerIleMetSerGluAsnArgSerArgGluGluLeuGlyValGlySerGlnSer 678
QY 1966 AGCTTTGCGGCGAGCATGGAATCATTTGAGGCTTCC 2001
DB SerPheSerGlySerMetGluIleIleGluValSer 690

RESULT 16
US-10-072-012-258
Sequence 258, Application US/10072012
Publication No. US20040033493A1
GENERAL INFORMATION:
APPLICANT: Tchernov, Vellizar
APPLICANT: Spytek, Kimberly
APPLICANT: Zernusen, Bryan
APPLICANT: Patutajan, Meera
APPLICANT: Shinkens, Richard
APPLICANT: Li, Li
APPLICANT: Gangolli, Esna
APPLICANT: Padigar, Muralidhara
APPLICANT: Anderson, David W.
APPLICANT: Rastelli, Luca
APPLICANT: Miller, Charles E.
APPLICANT: Gerlach, Valerie
APPLICANT: Taupier Jr, Raymond J.
APPLICANT: Gusev, Vladimir Y.
APPLICANT: Coleman, Steven D.
APPLICANT: Wolenc, Adam R.
APPLICANT: Pena, Carol E. A
APPLICANT: Furtak, Katarzyna
APPLICANT: Grose, William M.
APPLICANT: Alsbrook II, John P.
APPLICANT: Lepley, Denise M.
APPLICANT: Rieger, Daniel K.
APPLICANT: Burgess, Catherine E.
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-258
CURRENT APPLICATION NUMBER: US/10/072,012
PRIOR FILING DATE: 2002-01-31
PRIOR APPLICATION NUMBER: 60/265,102
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: 60/265,514
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,517
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,412
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,395
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/266,406
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 60/266,767

```

; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 258
; LENGTH: 662
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-012-258

Alignment Scores:
Pred. No.: 0 Length: 662
Score: 400.00 Matches: 469
Percent Similarity: 98.12% Conservative: 0
Best Local Similarity: 98.12% Mismatches: 0
Query Match: 59.97% Indels: 9
DB: 12 Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x US-10-072-012-258 (1-662)
QY 586 ACCGTGTCACAAAGCTGATCTTATCCCGAGTCTCATTTCTCGGTGCTGCTGATGAC 645
DB 194 ThCysProlyspProapPheileProgluSerHisPheleuArgValProValAsnAsp 213
QY 646 AGCTTTGTGAGAAATTTTGGCGGTGGACAATCAGATGATTGATTGAGAAAGCA 705
DB 214 SerPheCysGluValIleLeuProTyrLeuAspLysSerValAspPheileGluValAsa 233
QY 706 AAAGCCTCAATGATGTGTCTAGTGCACCTGTTAGCTGGATCTCCGCTCGCCACC 765
DB 234 LysAlaSerAsnGlyCysValIleuValHisCysLeuAlaGlyIleSerArgSerAlaThr 253
QY 766 ATGGCTATCCGCTACATCAGAGAGATGACATGCTCTTAAATGAAAGCTTACAGATT 825
DB 254 IleAlaIleAlaIleTyrIleMetLysArgMetAspMetSerLeuAspGluAlaTyrArgPhe 273
QY 826 GTGAAAGAAAAGAACCTATATCTCCAAATCTTCAATTTTCTGGGCAACCTCTGGAC 885
DB 274 ValLysGluValArgProThrIleSerProAsnPheAsnPheGluValLeuLeuAsp 293
QY 886 TATGAGAAAGATTTAAGAACAGACTGAGCATCAGAGGCCAAAGAGCAAACTCAAGCTG 945
DB 294 TyrGluLysLysIleLysAsnGlnThrGlyAlaSerGlyProLysSerLysLeuLysLeu 313
QY 946 CTGCACCTGAGAGAACCAATGAACTGTCTGCTGCTCTCAGAGGGTGAAGAAAGC 1005
DB 314 LeuHisIleuGluLysProAsnGluProValProAlaValSerGluGlyGlnLysSer 333
QY 1006 GAGAGCGCCCTCAGTCACCGCTGCGGCACTGCTCACTCAGAGGAGAGGCAAGG 1065
DB 334 GlnThrProLeuSerProProCysAlaAspSerAlaThrSerGluAlaAlaGlyGlnArg 353
QY 1066 CCGGTGATCCCGCAGCAGCTGCCAGCGTCCCGCAGCTGAGCGCTGTTAGAGAC 1125
DB 354 ProValHisProAla-----SerValProSerAlaGlnProSerLeuLeuGluAsp 370
QY 1126 AGCCCGCTGGTACAGGGGCTCAGTGGGCTGACCTGTCCGACAGACAGGCTGAAAGACG 1185
DB 371 SerProLeuValGlnAlaLeuSerGlyLeuHisIleuSerAlaAspArgLeuGluAspSer 390
QY 1186 AATAAGCTCAAGGCTCTCTCTCTGATATCAAAATCAGTTTCATTTGACGACAGATG 1245
DB 391 AsnLysLeuLysValArgSerPheSerLeuAspIleLysSerValSerLysAlaSerMet 410
QY 1246 GCAAGCATCTTACATGAGCTTCTCTCATCAGAGATGCTTGGATATCTACAAACCTTCC 1305
DB 411 AlaAlaSerLeuHisGlyPheSerSerSerGluAspAlaLeuGluLysTyrLysProSer 430

```

```

QY 1306 ACTACTGTGATGGAGCAACAAGTATGACCACTTCTCCCTGTTCAGAACTATCGAG 1365
DB 431 ThrThrLeuAspGlyThrAsnLysLeuCysGlnPheSerProValGlnGluLeuSerGlu 450
QY 1366 CAGACTCCCGAAAACAGCTCTGTATGAAGAGAGAGCCAGCATCCCAAGAAAGTGCAGACC 1425
DB 451 GlnThrProGluThrSerProAspLysGluGluAlaSerIleProLysLysLeuGlnThr 470
QY 1426 GCCAGGCTTCAGACAGCCAGAGCAAGCATTTGATTCGGTCAAGAACCCAGACAGTGGC 1485
DB 471 AlaArgProSerAspSerGlnSerLysArgLeuHisSerValArgThrSerSerSerGly 490
QY 1486 ACCGCCAGAGGCTCCCTTTATCTCCAGCTGACATGAAAGTGGAGCGGTGAGACAATTAC 1545
DB 491 ThrAlaGlnArgSerLeuLeuSerProLeuHisArgSerGlySerValGluAspAsnThr 510
QY 1546 CACACAGACTTCCTTTTCGCGCTTTCCACAGCCAGCAGCAGCTCAAGAGTGTGTGC 1605
DB 511 HisThrSerPheLeuPheGlyLeuSerThrSerGlnGlnHisIleuThrLysSerAlaGly 530
QY 1606 CTGGGCTTTAAGGGCTGGGCACTGGATATCTTGGCCCCCAGACCTTACCTTCCCTG 1665
DB 531 LeuGlyLeuLysGlyTyrPheIleSerAspIleLeuAlaProGlnThrSerThrProSerLeu 550
QY 1666 ACCAGAGCTGATTTTGGCAAGAGTCCCTACACTTCACTGCTGCTGAGCATCTAC 1725
DB 551 ThrSerSerTyrPheAlaThrGlnLysSerSerHisPheLysSerValAspAlaIleTyr 570
QY 1726 GGAGCAGTGCAGTACTCTGCTCAGCTGACAGCTGACAGCTGCCCACTTGGAGAGCA 1785
DB 571 GlyGlySerAlaSerLysIleSerAlaTyrSerCysSerGlnLeuProThrCysGlyAspGln 590
QY 1786 GTCTATTCTGTGGCGCAGGGCGGAGAGCCAGAGTGAACAGACTGCTGCGGCGAGCTGG 1845
DB 591 ValTyrSerValArgArgArgGlnLysProSerAspArgAlaAspSerArgArgSerTyr 610
QY 1846 CATGAGAGAGAGCCCTTTGAAAAGCAGTTTAAACGAGAGAGCTCCAAATGAATTTGA 1905
DB 611 HisGluGlnSerProGluGluLysGlnPheLysArgArgSerCysGlnMetGluPheGly 630
QY 1906 GAGAGCATATGTCAGAGAAACAGTCAAGGAGAGAGCTGGGGAAGTGGGCACTGCT 1965
DB 631 GluSerIleMetSerGluAsnArgSerArgGluGluLeuGlyLysValGlySerGlnSer 650
QY 1966 AGCTTTTCGGGCAAGATGAAATCATTGAGCTTCC 2001
DB 651 SerPheSerGlySerMetGluIleIleGluValSer 662

RESULT 17
US-10-072-012-256
; Sequence 256, Application US/10072012
; Publication No. US2004003493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Speitek, Kimberly
; APPLICANT: Zernhusen, Bryan
; APPLICANT: Patuturajan, Meera
; APPLICANT: Shimkete, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangolli, Beha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Coleman, Steven D.
; APPLICANT: Molenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grose, William M.

```

```

/ APPLICANT: Alsobrook II, John P.
/ APPLICANT: Lepley, Denise M.
/ APPLICANT: Rieger, Daniel K.
/ APPLICANT: Burgess, Catherine E.
/ TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
/ FILE REFERENCE: 21402-258
/ CURRENT APPLICATION NUMBER: US/10/072,012
/ PRIOR FILING DATE: 2002-01-31
/ PRIOR APPLICATION NUMBER: 60/265,102
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: 60/265,514
/ PRIOR FILING DATE: 2001-01-31
/ PRIOR APPLICATION NUMBER: 60/265,517
/ PRIOR FILING DATE: 2001-01-31
/ PRIOR APPLICATION NUMBER: 60/265,412
/ PRIOR FILING DATE: 2001-01-31
/ PRIOR APPLICATION NUMBER: 60/265,395
/ PRIOR FILING DATE: 2001-01-31
/ PRIOR APPLICATION NUMBER: 60/266,406
/ PRIOR FILING DATE: 2001-02-02
/ PRIOR APPLICATION NUMBER: 60/266,767
/ PRIOR FILING DATE: 2001-02-05
/ PRIOR APPLICATION NUMBER: 60/267,057
/ PRIOR FILING DATE: 2001-02-07
/ PRIOR APPLICATION NUMBER: 60/266,975
/ PRIOR FILING DATE: 2001-02-07
/ PRIOR APPLICATION NUMBER: 60/267,459
/ PRIOR FILING DATE: 2001-02-08
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 1391
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 256
/ LENGTH: 680
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-072-012-256

Alignment Scores:
Pred. No.: 0 Length: 680
Score: 394.00 Matches: 394
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 59.07% Indels: 0
DB: 12 Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x US-10-072-012-256 (1-680)
QY 820 AGATTGTGAAGAAAAAGACCTACATATCTCAAACTTCAATTTTCTGGCCCAACTC 879
DB 287 ArgPheValIysGluYsaIgrProIhTlIeSerProAsnPhenAsnPhenLeuGlyGlnLeu 306
QY 880 CTGACATATGAGAAAGAAAGATTAAAGAACAGATGAGCATCAGGGCCAAAGAACAACTC 939
DB 307 LeuAspTyrGluIysIysIleYsaGlnThrGlyAlaSerGlyProIysSerIysLeu 326
QY 940 AAGCTGTGACCTGAGAAAGCAATGAACCTGCTCCCTGCTCTCAGAGGCTGACAG 999
DB 327 LysLeuLeuHisLeuGluYsaIgrProAsnGlnProValProAlaValSerGlnIlyGln 346
QY 1000 AAAAGCGAGACGCCCTCACTGACCACTGTCGCACTGTGTAACCTGACGAGGACAGAGA 1059
DB 347 LysSerGluThrProIysSerProProCysAlaAspSerAlaThrSerGlnAlaIleGly 366
QY 1060 CAAAGCCCGTGATCCCGCAGCGCTGCCAGCGTCCAGCGTCCAGCGTCCAGCGCTGTTA 1119
DB 367 GlnArgProValHisProAlaSerValProSerValProSerValGlnProSerLeuLeu 386
QY 1120 GAGGACAGCCCGGTGTATAGCGGCTCAGTGGGCTGCACCTCTCCGACAGACAGGCTGAA 1179
DB 387 GluAspSerProLeuValGlnAlaLeuSerGlyLeuHisLeuSerAlaAspArgLeuGln 406
QY 1180 GACAGCAATTAAGCTCAAGCGTCTCTCTCTGATATCAATCATGTTCAATATTCAGGC 1239
DB 1180 GACAGCAATTAAGCTCAAGCGTCTCTCTCTGATATCAATCATGTTCAATATTCAGGC 1239
```

```

DB 407 AspSerAsnIysLeuYsaIgrSerPheSerLeuAspIleYsaSerValSerTyrSerAla 426
QY 1240 AGCATGGACAGACCTCTTACATGGCTCTCTCTCATCAGAAAGATGTTGGAAATACACAA 1299
DB 427 SerMetAlaIaSerLeuHisGlyPheSerSerSerSerSerSerSerSerSerSerSerSer 446
QY 1300 CCTTCCACATCTCTGATGGAGCAACAAAGCATATGCACTTCTCCCTGTTGAGAACTA 1359
DB 447 ProSerThrThrLeuAspGlyThrAsnIysLeuYsaIgrPheSerProValGlnIleuLeu 466
QY 1360 TCGGACAGACCTCCGAAACCAAGTCCTGATAGAGAGGAAGCCAGCATCCCAAGACCTG 1419
DB 467 SerGlnGlnThrProGlnIysSerProAsnIysGlnGlnAlaSerIleProIysIysLeu 486
QY 1420 CAGACCGCAGGCGCTTACAGACCCAGACCAAGCATGCAATTCGTCAGAACCGACGAC 1479
DB 487 GlnThrAlaArgProSerAspSerSerGlnSerIysAlaGlnLeuHisSerValArgThrSer 506
QY 1480 AGTGGACCGCCCAAGAGTCCCTTTATCTCACTGCATCGAAGTGGAGCGTGGAGAC 1539
DB 507 SerGlyThrAlaGlnArgSerLeuLeuSerProLeuHisArgSerGlySerValGlnAsp 526
QY 1540 AATTACCAACACAGCTCTCTTTCGACCTTTCACACAGCAGACGACCTCACAAAGCT 1599
DB 527 AsnTyrHisThrSerPheLeuPheGlyLeuSerThrSerGlnGlnHisLeuThrLysSer 546
QY 1600 GCTGCGCTGGGCTTAAAGGCTGAGCATCGAATATCTTGAGCCCGCCAGACCTTACCCCT 1659
DB 547 AlaGlyLeuGlyLeuYsaIgrTrpHisSerAspIleLeuAlaProGlnThrSerThrPro 566
QY 1660 TCCCTGACAGAGCGTGTATTTTCCACAGAGTCCACACCTTCTACTGCTCTCAGCC 1719
DB 567 SerLeuThrSerSerThrPyrPheAlaThrIysSerThrHisPheTyrSerAlaSerAla 586
QY 1720 ATCTAAGGAGGACAGTCCAGTACTCTGCTCCTACAGTGCAGCCAGCTGCCACTTGCGGA 1779
DB 587 IleTyrGlyGlySerAlaSerTyrSerAlaTyrSerIysSerSerGlnLeuProIhTnCysGly 606
QY 1780 GACCAAGCTATTTCTGTGCGGACGCGGCAAGAACCAAGTGAACAGAGTGACTGCGGCGG 1839
DB 607 AspGlnValTyrSerValArgArgArgGlnIysProSerAspArgAlaAspSerArgArg 626
QY 1840 AGCTGCGATGAAGAGAGCCCTTGAAGAGAGATTAAAGCGAGCAAGCTGCCAAATGGA 1899
DB 627 SerTrpHisGlnIysSerProPheGlnIysGlnPheYsaIgrArgSerCysGlnMetGln 646
QY 1900 TTGGAAGAGCATCATGTATGAGAAACAGGTCAACGGAAGAGCTGGGGAAAGTGGCGAGT 1959
DB 647 PheGlyIysSerIleMetSerGlnAsnArgSerArgGlnIleuGlyValIysValIysSer 666
QY 1960 CAGTCTAGCTTTTCGGGACAGCATGGAATCATTTGAGGTCTCC 2001
DB 667 GlnSerSerPheSerGlySerMetGlnIleIleGlnValSer 680

RESULT 18
US-10-296-115-1259
/ Sequence 1259, Application US/10296115
/ Publication No. US20040053248A1
/ GENERAL INFORMATION:
/ APPLICANT: Hysq Inc
/ TITLE OF INVENTION: No. US20040053248A1 Nucleic Acids and Polypeptides
/ FILE REFERENCE: 784PCT
/ CURRENT APPLICATION NUMBER: US/10/296,115
/ CURRENT FILING DATE: 2002-11-18
/ PRIOR APPLICATION NUMBER: US09/488,725
/ PRIOR FILING DATE: 2000-01-21
/ PRIOR APPLICATION NUMBER: US09/552,317
/ PRIOR FILING DATE: 2000-04-25
/ NUMBER OF SEQ ID NOS: 1478
/ SEQ ID NO 1259
/ LENGTH: 672
/ TYPE: PRT
/ ORGANISM: Homo sapiens
```

```
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)-(672)
OTHER INFORMATION: xaa = any amino acid or other as shown in Table 3
US-10-296-115-1259

Alignment Scores:
Pred. No.: 2,556-238 Length: 672
Score: 258.00 Matches: 258
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 38,68% Indels: 0
DB: 12 Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x US-10-296-115-1259 (1-672)

QY 586 ACCGTGTCAGAAAGCTGATTTATCCCGAGTCTCATTTCTGCGTGTGCTGTAATGAC 645
DB 201 ThrcyProlyserProapPheileProgluserHibheleuargValProvalaamdp 220

QY 646 AGCTTTTGAGAAATTTTGGCCGTGGTGAGCAATCAGTATTCATTGAGAAAGCA 705
DB 221 SerPheCysglulysleuProtyrleuapblyseValapbelleglulysAla 240

QY 706 AAAGCTCCAGATGATGTTCTAGTGCACCTGTTAGCTGGATCTCCGCTCCGACACC 765
DB 241 LysAlaSeramndlyCysValleuValHisCysleuAlaaglylleSeratgserAlatm 260

QY 766 ATGCTATCGCTCATCATGAGAGAGATGAGACATGCTTTAGATGAGCTTACAGATT 825
DB 261 IleAlaIlelatIrylleMetlysaIrgMetCaspMeSerleuapbglulalatyrrargphe 280

QY 826 GTGAAGAAAAAAGACTACTATATCTCCAACTTCAATTTTCTGGCCCAACTCTCGAC 885
DB 281 ValLysglulysarIprorHrIleSerProamPheamPheleuglylneuleuamdp 300

QY 886 TATGAGAAGAAATTAAAGACAGACAGATGAGATGAGGAGCAAGAACTCAAGCTG 945
DB 301 TyrgIulysleyleysahengIntHrglyAlaSerClyProlyserlyleuylsleu 320

QY 946 CTGACCTGAGAGAGCCAAATGAACCTGCTCTGCTGTCTCAGAGGGTGAGCAAGAAAG 1005
DB 321 LeuHileugIulysrProaengIuprovalProalavalsergluglylneulysSer 340

QY 1006 GAAAGCCCTCACTCACTCCAGCTGTGCGGACTCTGCTACCTCAAGGAGCAAGCAAG 1065
DB 341 GlutHrProleuserProProCysAlaAapSerAlatHrserglulalaglylndary 360

QY 1066 CCCGTCATCCCGCCAGCGTCCAGCGTCCAGCGTCCAGCGCTGTTAGAGGAC 1125
DB 361 ProvalHrProAlaSerValProserValProserValInProserleuLeuGlulap 380

QY 1126 AGCCCGCTGTAGAGCGCTCAGTGGGCTGACCTGTCCGACAGACAGCTGAGAGAC 1185
DB 381 SerProleuValglinalaleuserglyleuHileuserlalaapbIrgleuGlulapSer 400

QY 1186 AATAAGCTCAAGGCTTCTCTCTGAGATATCAATCACTTTCATATTCAGCCAGCAT 1245
DB 401 AsnlyleuylsarIrgserPheSerleuapbIlelyseValaIserlyrserAlaSerMet 420

QY 1246 GCAGCATCTTATAGGCTCTGCTCATCAGAGAGCTTGAATAATCAAGCAAGCTTCC 1305
DB 421 AlaAlaSerleuHilegllyPheSerSerSerGlulapbAlaLeuGlulatyrrlyrProser 440

QY 1306 ACTACTCTGGATGAGCAAGCAAGCTATGCGCAGTCTCCCTGTTCAAGAACTA 1359
DB 441 ThrThrleuapbgllyThrAenlyleuCyglInPheSerProvalglInglulau 458

RESULT 19
US-10-072-012-682
; Sequence 682, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:
```

```
APPLICANT: Tchernev, Velizar
APPLICANT: Spytek, Kimberly
APPLICANT: Zerhusen, Bryan
APPLICANT: Paturajan, Meera
APPLICANT: Shimkete, Richard
APPLICANT: Li, Li
APPLICANT: Gangolli, Bha
APPLICANT: Padigaru, Muralidhara
APPLICANT: Anderson, David W.
APPLICANT: Rastelli, Luca
APPLICANT: Miller, Charles E.
APPLICANT: Gerlach, Valerie
APPLICANT: Taupier Jr, Raymond J.
APPLICANT: Gusev, Vladimr Y.
APPLICANT: Coleman, Steven D.
APPLICANT: Wolenc, Adam R.
APPLICANT: Pena, Carol E. A
APPLICANT: Furtak, Katarzyna
APPLICANT: Grose, William M.
APPLICANT: Alsobrook II, John P.
APPLICANT: Lepley, Denise M.
APPLICANT: Rieger, Daniel K.
APPLICANT: Burgess, Catherine E.
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-258
CURRENT APPLICATION NUMBER: US/10/072,012
CURRENT FILING DATE: 2002-01-31
PRIOR APPLICATION NUMBER: 60/265,102
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: 60/265,514
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,517
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,412
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,395
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/266,406
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 60/266,767
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: 60/267,057
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/266,975
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/267,459
PRIOR FILING DATE: 2001-02-08
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 1391
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 682
LENGTH: 660
TYPE: PRT
ORGANISM: Mus musculus
US-10-072-012-682

Alignment Scores:
Pred. No.: 6,486-55 Length: 660
Score: 67.00 Matches: 67
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 10,04% Indels: 0
DB: 12 Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x US-10-072-012-682 (1-660)

QY 49 TTGGTGGCTCTGCTGGAAGGAGCAAGAAAGTGCTGCTATGATGCGCGCATTT 108
DB 15 LeuValAlaLeuLeuLeuIsergllyHrIrglyllyValleuLeuIleapSerHrgrProPhe 34

QY 109 GTGATATCAATATACATCCACATTTTGAAGCCATTATATCAATGCTTCAAGCTTATG 168
DB 35 ValglulTyraenThrserHrIleleugIulalIleamIleamCysSerlyleuMet 54
```


QY 169 AAGCGAAGTTGCAACAGACAAAGTTAATTAACAGAGCTCATCCAGCATTGCGGAAA 228
 |||||
 Db 55 LysArgGArgLeuGInGInAAsPlySValLeuLeIthrGInuLeuIleGInHISerAlaLys 74
 |||||
 QY 229 CATAAAGTTGACATTGATTGC 249
 |||||
 Db 75 HisLysValAAsPlyIleAsPlys 81
 |||||
 RESULT 20
 US-10-072-012-683
 ; Sequence 683, Application US/10072012
 ; Publication No. US20040033493A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Tchernev, Velizar
 ; APPLICANT: Spytek, Kimberly
 ; APPLICANT: Zernhusen, Bryan
 ; APPLICANT: Patturajan, Meera
 ; APPLICANT: Shinkets, Richard
 ; APPLICANT: Li, Li
 ; APPLICANT: Gangolli, Esha
 ; APPLICANT: Padigaru, Muralidhara
 ; APPLICANT: Anderson, David W.
 ; APPLICANT: Rastelli, Luca
 ; APPLICANT: Miller, Charles E.
 ; APPLICANT: Gerlach, Valerie
 ; APPLICANT: Taupier Jr, Raymond J.
 ; APPLICANT: Gusev, Vladimir Y.
 ; APPLICANT: Coleman, Steven D.
 ; APPLICANT: Wolenc, Adam R.
 ; APPLICANT: Pena, Carol E. A
 ; APPLICANT: Futak, Katarzyna
 ; APPLICANT: Grose, William M.
 ; APPLICANT: Alsobrook II, John P.
 ; APPLICANT: Lepley, Denise M.
 ; APPLICANT: Rieger, Daniel K.
 ; APPLICANT: Burgess, Catherine E.
 ; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
 ; FILE REFERENCE: 21402-258
 ; CURRENT APPLICATION NUMBER: US/10/072,012
 ; CURRENT FILING DATE: 2002-01-31
 ; PRIOR APPLICATION NUMBER: 60/265,102
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: 60/265,514
 ; PRIOR FILING DATE: 2001-01-31
 ; PRIOR APPLICATION NUMBER: 60/265,517
 ; PRIOR FILING DATE: 2001-01-31
 ; PRIOR APPLICATION NUMBER: 60/265,412
 ; PRIOR FILING DATE: 2001-01-31
 ; PRIOR APPLICATION NUMBER: 60/265,395
 ; PRIOR FILING DATE: 2001-01-31
 ; PRIOR APPLICATION NUMBER: 60/266,406
 ; PRIOR FILING DATE: 2001-02-02
 ; PRIOR APPLICATION NUMBER: 60/266,767
 ; PRIOR FILING DATE: 2001-02-05
 ; PRIOR APPLICATION NUMBER: 60/267,057
 ; PRIOR FILING DATE: 2001-02-07
 ; PRIOR APPLICATION NUMBER: 60/266,975
 ; PRIOR FILING DATE: 2001-02-07
 ; PRIOR APPLICATION NUMBER: 60/267,459
 ; PRIOR FILING DATE: 2001-02-08
 ; Remaining Prior Application data removed - See File Wrapper or PALM.
 ; NUMBER OF SEQ ID NOS: 1391
 ; SOFTWARE: PatencIn Ver. 2.1
 ; SEQ ID NO 683
 ; LENGTH: 677
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 ; US-10-072-012-683
 Alignment Scores: 4,91e-52 Length: 677
 Pred. No.: 64.00 Matches: 64
 Score:

Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 9.60% Indels: 0
 Db: 12 Gaps: 0
 US-10-029-345a-108_COPY_532_2532 (1-2001) x US-10-072-012-683 (1-677)
 QY 346 AGTTCACTCTGTTTCACCTGCTTGACAGGTGGGTTTGTGAGTTCTTCGTGTTCCCT 405
 |||||
 Db 114 SerPheAsnSerValHisLeuLeuAlaGlyGlyPheAlaGluPheSerArgCysPhePro 133
 |||||
 QY 406 GGCCTCTGTGAAGAAAATCCACTAGCCCTACCTGCATTTCACGCTTGTACT 465
 |||||
 Db 134 GlyLeuGysGInGlyLysSerThrLeuValProThrCysIleSerGlnProCysLeuPro 153
 |||||
 QY 466 GTTGCACATTGGGCCAACCCGAAATTCCTCCATCTTATTCATGCGTCGACGAGAT 525
 |||||
 Db 154 ValAlaAsnIleGlyProThrArgIleLeuProAsnLeuTyrlLeuGlyCysGlnArgasp 173
 |||||
 QY 526 GTTCCACACAG 537
 |||||
 Db 174 ValLeuAsnLys 177
 |||||
 RESULT 21
 US-10-072-012-702
 ; Sequence 702, Application US/10072012
 ; Publication No. US20040033493A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Tchernev, Velizar
 ; APPLICANT: Spytek, Kimberly
 ; APPLICANT: Zernhusen, Bryan
 ; APPLICANT: Patturajan, Meera
 ; APPLICANT: Shinkets, Richard
 ; APPLICANT: Li, Li
 ; APPLICANT: Gangolli, Esha
 ; APPLICANT: Padigaru, Muralidhara
 ; APPLICANT: Anderson, David W.
 ; APPLICANT: Rastelli, Luca
 ; APPLICANT: Miller, Charles E.
 ; APPLICANT: Gerlach, Valerie
 ; APPLICANT: Taupier Jr, Raymond J.
 ; APPLICANT: Gusev, Vladimir Y.
 ; APPLICANT: Coleman, Steven D.
 ; APPLICANT: Wolenc, Adam R.
 ; APPLICANT: Pena, Carol E. A
 ; APPLICANT: Futak, Katarzyna
 ; APPLICANT: Grose, William M.
 ; APPLICANT: Alsobrook II, John P.
 ; APPLICANT: Lepley, Denise M.
 ; APPLICANT: Rieger, Daniel K.
 ; APPLICANT: Burgess, Catherine E.
 ; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
 ; FILE REFERENCE: 21402-258
 ; CURRENT APPLICATION NUMBER: US/10/072,012
 ; CURRENT FILING DATE: 2002-01-31
 ; PRIOR APPLICATION NUMBER: 60/265,102
 ; PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: 60/265,514
 ; PRIOR FILING DATE: 2001-01-31
 ; PRIOR APPLICATION NUMBER: 60/265,517
 ; PRIOR FILING DATE: 2001-01-31
 ; PRIOR APPLICATION NUMBER: 60/265,412
 ; PRIOR FILING DATE: 2001-01-31
 ; PRIOR APPLICATION NUMBER: 60/265,395
 ; PRIOR FILING DATE: 2001-01-31
 ; PRIOR APPLICATION NUMBER: 60/266,406
 ; PRIOR FILING DATE: 2001-02-02
 ; PRIOR APPLICATION NUMBER: 60/266,767
 ; PRIOR FILING DATE: 2001-02-05
 ; PRIOR APPLICATION NUMBER: 60/267,057
 ; PRIOR FILING DATE: 2001-02-07
 ; PRIOR APPLICATION NUMBER: 60/266,975
 ; PRIOR FILING DATE: 2001-02-07


```

; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 702
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (469)
; OTHER INFORMATION: Wherein Xaa is any amino acid.
US-10-072-012-702

Alignment Scores:
Pred. No.: 9.03e-10 Length: 501
Score: 20.00 Matches: 20
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.00% Indels: 0
DB: 12 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-10-072-012-702 (1-501)

QY 730 GTCGACTGTTTACGTGGATCTCCGCTCCGCCACCATGCTATCGCCTACATCATGAG 789
DB 98 ValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIleMetLys 117

RESULT 22
US-10-072-012-699
; Sequence 699, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zernusen, Bryan
; APPLICANT: Paturajan, Meera
; APPLICANT: Shimkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangoli, Bsha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Coleman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grose, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
```

```

; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459.
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 699
; LENGTH: 625
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-012-699

Alignment Scores:
Pred. No.: 8.81e-10 Length: 625
Score: 20.00 Matches: 20
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.00% Indels: 0
DB: 12 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-10-072-012-699 (1-625)

QY 730 GTCGACTGTTTACGTGGATCTCCGCTCCGCCACCATGCTATCGCCTACATCATGAG 789
DB 244 ValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIleMetLys 263

RESULT 23
US-10-072-012-700
; Sequence 700, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zernusen, Bryan
; APPLICANT: Paturajan, Meera
; APPLICANT: Shimkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangoli, Bsha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Coleman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grose, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
```

RESULT 24
 US-10-072-012-266
 Sequence 266, Application US/10072012
 Publication No. US2004003493A1
 GENERAL INFORMATION:
 APPLICANT: Tchernen, Velizar
 APPLICANT: Spytek, Kimberly
 APPLICANT: Zernusen, Bryan
 APPLICANT: Patnurajan, Meera
 APPLICANT: Shinkets, Richard
 APPLICANT: Li, Li
 APPLICANT: Gangolli, Esha
 APPLICANT: Padigar, Muralidhara
 APPLICANT: Anderson, David W.
 APPLICANT: Rastelli, Luca
 APPLICANT: Miller, Charles B.
 APPLICANT: Gerlach, Valerie
 APPLICANT: Taupler Jr, Raymond J.
 APPLICANT: Gusev, Vladimir Y.
 APPLICANT: Colman, Steven D.
 APPLICANT: Wolenc, Adam R.
 APPLICANT: Pena, Carol E. A
 APPLICANT: Furtak, Katarzyna
 APPLICANT: Grosse, William M.
 APPLICANT: Alsobrook II, John P.
 APPLICANT: Lepley, Denise M.
 APPLICANT: Rieger, Daniel K.
 APPLICANT: Burgess, Catherine E.
 TITLE OF INVENTION: proteins and Nucleic Acids Encoding Same
 FILE REFERENCE: 21402-258
 CURRENT APPLICATION NUMBER: US/10/072, 012
 CURRENT FILING DATE: 2002-01-31
 PRIOR APPLICATION NUMBER: 60/265, 102
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: 60/265, 514
 PRIOR FILING DATE: 2001-01-31
 PRIOR APPLICATION NUMBER: 60/265, 517
 PRIOR FILING DATE: 2001-01-31
 PRIOR APPLICATION NUMBER: 60/265, 412
 PRIOR FILING DATE: 2001-01-31

```

RESULT 25
US-10-282-122A-57268
/ Sequence 57268, Application US/10282122A
/ Publication No. US20040029129A1
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Liangsu
/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Malone, Cheryl
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Kai
/ APPLICANT: Zyskind, Judith
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John
/ APPLICANT: Carr, Grant
/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Forsyth, R.
/ APPLICANT: Xu, H.
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: ELITPA.034A
/ CURRENT APPLICATION NUMBER: US/10/282,122A
/ CURRENT FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636

```

```
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 57268
/ LENGTH: 498
/ TYPE: PRT
/ ORGANISM: Enterococcus faecalis
US-10-282-122A-57268

Alignment Scores:
Pred. No.: 3 62      Length: 498
Score: 10 00      Matches: 10
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 1.55%      Indels: 0
DB: 12      Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-10-282-122A-57268 (1-498)
QY 103 GCGCGCTATCAATTAGCAGCACTTTTCCG 74
Db 342 AlagLYTcgtlnhnuhlaiaaleupnePro 351

RESULT 26
US-10-225-060-11
/ Sequence 11, Application US/10225060
/ Publication No. US20030092891A1
/ GENERAL INFORMATION:
/ APPLICANT: Davis et al.
/ TITLE OF INVENTION: Expressed Ligand - Vascular Intercellular Signaling
/ FILE REFERENCE: REG 333-Z
/ CURRENT APPLICATION NUMBER: US/10/225,060
/ PRIOR FILING DATE: 2002-08-21
/ PRIOR APPLICATION NUMBER: US/09/709,188
/ PRIOR FILING DATE: 2000-11-09
/ PRIOR FILING DATE: 1996-10-25
/ NUMBER OF SEQ ID NOS: 30
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 11
/ LENGTH: 503
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-225-060-11

Alignment Scores:
Pred. No.: 301      Length: 503
Score: 8 00      Matches: 8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 1.20%      Indels: 0
DB: 14      Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-10-225-060-11 (1-503)
QY 1083 CGTGCCGAGCGTGCCGCGTGCA 1106
Db 80 ArgAlaGlnArgAlaGlnArgAla 87

RESULT 27
US-10-214-812-2
/ Sequence 2, Application US/10214812
/ Publication No. US20030064470A1
/ GENERAL INFORMATION:
/ APPLICANT: Valenzuela et al.
/ TITLE OF INVENTION: NOVEL LIGANDS, METHODS OF MAKING AND USES THEREOF
/ FILE REFERENCE: REG330-K
/ CURRENT APPLICATION NUMBER: US/10/214,812
/ CURRENT FILING DATE: 2002-08-08
```

```
/ PRIOR APPLICATION NUMBER: US/09/202,491
/ PRIOR FILING DATE: 1998-11-16
/ PRIOR APPLICATION NUMBER: PCT/US97/10728
/ PRIOR FILING DATE: 1997-06-19
/ PRIOR APPLICATION NUMBER: 60/022,999
/ PRIOR FILING DATE: 1996-08-02
/ PRIOR APPLICATION NUMBER: 60/021,087
/ PRIOR FILING DATE: 1996-07-02
/ PRIOR APPLICATION NUMBER: 08/665,926
/ PRIOR FILING DATE: 1996-06-19
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 2
/ LENGTH: 509
/ TYPE: PRT
/ ORGANISM: Mus musculus
US-10-214-812-2

Alignment Scores:
Pred. No.: 300      Length: 509
Score: 8 00      Matches: 8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 1.20%      Indels: 0
DB: 12      Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-10-214-812-2 (1-509)
QY 1083 CGTGCCGAGCGTGCCGCGTGCA 1106
Db 86 ArgAlaGlnArgAlaGlnArgAla 93

RESULT 28
US-10-214-812-3
/ Sequence 3, Application US/10214812
/ Publication No. US20030064470A1
/ GENERAL INFORMATION:
/ APPLICANT: Valenzuela et al.
/ TITLE OF INVENTION: NOVEL LIGANDS, METHODS OF MAKING AND USES THEREOF
/ FILE REFERENCE: REG330-K
/ CURRENT APPLICATION NUMBER: US/10/214,812
/ PRIOR FILING DATE: 2002-08-08
/ PRIOR APPLICATION NUMBER: US/09/202,491
/ PRIOR FILING DATE: 1998-11-16
/ PRIOR APPLICATION NUMBER: PCT/US97/10728
/ PRIOR FILING DATE: 1997-06-19
/ PRIOR APPLICATION NUMBER: 60/022,999
/ PRIOR FILING DATE: 1996-08-02
/ PRIOR APPLICATION NUMBER: 60/021,087
/ PRIOR FILING DATE: 1996-07-02
/ PRIOR APPLICATION NUMBER: 08/665,926
/ PRIOR FILING DATE: 1996-06-19
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 3
/ LENGTH: 509
/ TYPE: PRT
/ ORGANISM: Mus musculus
US-10-214-812-3

Alignment Scores:
Pred. No.: 300      Length: 509
Score: 8 00      Matches: 8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 1.20%      Indels: 0
DB: 12      Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-10-214-812-3 (1-509)
QY 1083 CGTGCCGAGCGTGCCGCGTGCA 1106
Db 86 ArgAlaGlnArgAlaGlnArgAla 93
```

```
RESULT 29
US-10-215-224-2
; Sequence 2, Application US/10215224
; Publication No. US20030059887A1
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al.
; TITLE OF INVENTION: NOVEL LIGANDS, METHODS OF MAKING AND USES THEREOF
; FILE REFERENCE: REG330-K
; CURRENT APPLICATION NUMBER: US/10/215,224
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: PCT/US97/10728
; PRIOR FILING DATE: 1998-11-16
; PRIOR APPLICATION NUMBER: 60/021,087
; PRIOR FILING DATE: 1996-06-19
; PRIOR APPLICATION NUMBER: 08/665,926
; PRIOR FILING DATE: 1996-06-19
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-215-224-2

Alignment Scores:
Pred. No.: 300          Length: 509
Score: 8.00           Matches: 8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%    Mismatches: 0
Query Match: 1.20%              Indels: 0
DB: 14                     Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x US-10-215-224-2 (1-509)
QY 1083 CGTCCCGAGCGTCCCGAGCGTGCA 1106
Db 86 ArgAlaGlnArgAlaGlnArgAla 93

RESULT 30
US-10-215-224-3
; Sequence 3, Application US/10215224
; Publication No. US20030059887A1
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al.
; TITLE OF INVENTION: NOVEL LIGANDS, METHODS OF MAKING AND USES THEREOF
; FILE REFERENCE: REG330-K
; CURRENT APPLICATION NUMBER: US/10/215,224
; PRIOR FILING DATE: 2002-08-08
; PRIOR APPLICATION NUMBER: US/09/202,491
; PRIOR FILING DATE: 1998-11-16
; PRIOR APPLICATION NUMBER: PCT/US97/10728
; PRIOR FILING DATE: 1997-06-19
; PRIOR APPLICATION NUMBER: 60/022,999
; PRIOR FILING DATE: 1996-08-02
; PRIOR APPLICATION NUMBER: 60/021,087
; PRIOR FILING DATE: 1996-07-02
; PRIOR APPLICATION NUMBER: 08/665,926
; PRIOR FILING DATE: 1996-06-19
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-215-224-3

Alignment Scores:
Pred. No.: 300          Length: 509
Score: 8.00           Matches: 8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%    Mismatches: 0
Query Match: 1.20%              Indels: 0
DB: 14                     Gaps: 0
```

```
Score: 8.00          Matches: 8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%    Mismatches: 0
Query Match: 1.20%              Indels: 0
DB: 14                     Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x US-10-215-224-3 (1-509)
QY 1083 CGTCCCGAGCGTCCCGAGCGTGCA 1106
Db 86 ArgAlaGlnArgAlaGlnArgAla 93

RESULT 31
US-10-225-060-10
; Sequence 10, Application US/10225060
; Publication No. US20030092891A1
; GENERAL INFORMATION:
; APPLICANT: Davis et al.
; TITLE OF INVENTION: Expressed Ligand - Vascular Intercellular Signaling
; FILE REFERENCE: REG 333-Z
; CURRENT APPLICATION NUMBER: US/10/225,060
; PRIOR FILING DATE: 2002-08-21
; PRIOR APPLICATION NUMBER: US/09/709,188
; PRIOR FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: 08/740,223
; PRIOR FILING DATE: 1996-10-25
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-225-060-10

Alignment Scores:
Pred. No.: 300          Length: 509
Score: 8.00           Matches: 8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%    Mismatches: 0
Query Match: 1.20%              Indels: 0
DB: 14                     Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x US-10-225-060-10 (1-509)
QY 1083 CGTCCCGAGCGTCCCGAGCGTGCA 1106
Db 86 ArgAlaGlnArgAlaGlnArgAla 93

RESULT 32
US-09-833-745-57
; Sequence 57, Application US/09833745
; Patent No. US20020052038A1
; GENERAL INFORMATION:
; APPLICANT: ROBERTS, JOSEPH
; APPLICANT: SETHURAMAN, NARAJAN
; APPLICANT: MACALLISTER, THOMAS
; TITLE OF INVENTION: CLONING, OVEREXPRESSION AND THERAPEUTIC USE OF
; FILE REFERENCE: 078728/0106
; CURRENT APPLICATION NUMBER: US/09/833,745
; PRIOR FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/197,770
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 57
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-833-745-57

Alignment Scores:
```

Pred. No.: 300 Length: 513
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 9 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-09-833-745-57 (1-513)

QY 137 AGCCATTATATCACTGCTCCA 160
DB 459 LyePLeuLeuLeuSerThrAlaPro 466

RESULT 33
US-10-424-599-231545
; Sequence 231545, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223) B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 231545
; LENGTH: 529
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_51106C.1.pap
US-10-424-599-231545

Alignment Scores:
Pred. No.: 299 Length: 529
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 12 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-10-424-599-231545 (1-529)

QY 374 GTGGTTTGTGAGTCTCTCGTT 397
DB 512 ValGlyLeuLeuSerSerLeuVal 519

RESULT 34
US-08-485-163-5
; Sequence 5, Application US/08485163
; Publication No. US20020098191A1
; GENERAL INFORMATION:
; APPLICANT: Beaudry, Gary A.
; APPLICANT: Maddon, Paul J.
; TITLE OF INVENTION: CD4-GAMMA2 CD4-IGG2 CHIMERAS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,163

FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 37690-II-1-PCT-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 530 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: homo sapien
; CELL TYPE: lymphocyte
US-08-485-163-5

Alignment Scores:
Pred. No.: 299 Length: 530
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 0 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-08-485-163-5 (1-530)

QY 1819 GACAGAGCTGACTCGCGCGGAGC 1842
DB 78 AspArgAlaAspSerArgSer 85

RESULT 35
US-09-766-995-4
; Sequence 4, Application US/09766995
; Patent No. US20020052481A1
; GENERAL INFORMATION:
; APPLICANT: Graham P. Allaway et al.
; TITLE OF INVENTION: NON-PEPTIDYL MOIETY-CONJUGATED CD4-GAMMA2 AND CD4-IGG2 IMMUNOCON
; FILE REFERENCE: 2048/41215-CB/JPW/SHS
; CURRENT APPLICATION NUMBER: US/09/766,995
; CURRENT FILING DATE: 2001-01-22
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 4
; LENGTH: 530
; TYPE: PRT
; ORGANISM: homo sapians
US-09-766-995-4

Alignment Scores:
Pred. No.: 299 Length: 530
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 9 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-09-766-995-4 (1-530)

QY 1819 GACAGAGCTGACTCGCGCGGAGC 1842
DB 78 AspArgAlaAspSerArgSer 85

RESULT 36
US-09-939-537-6
; Sequence 6, Application US/09939537
; Publication No. US20030138410A1

GENERAL INFORMATION:
APPLICANT: Seed, Brian
Banapour, Babak
Romeo, Charles
Kolatus, Waldemar
TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
CELLS BY CHIMERIC CD4 RECEPTOR- BEARING CELLS
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Clark & Elbing LLP
STREET: 176 Federal Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: Fasteq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/939,537
FILING DATE: 24-Aug-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/284,391
FILING DATE: 02-AUG-1994
APPLICATION NUMBER: 08/195,395
FILING DATE: 14-FEB-1994
APPLICATION NUMBER: 07/847,566
FILING DATE: 06-MAR-1992
APPLICATION NUMBER: 07/665,961
FILING DATE: 07-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Elbing, Karen L.
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/247001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-939-537-6
Alignment Scores:
Pred. No.: 299 Length: 532
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 10 Gaps: 0
US-10-029-345A-108_COPY_532_2532 (1-2001) x US-09-939-537-6 (1-532)
QY 1819 GACGAGCTGACTGCGGCGGAGC 1842
DB 78 AspArgAlaAspSerArgArgSer 85
RESULT 37
US-09-243-008-6
Sequence 6, Application US/09243008
Publication No. US20040005334A1
GENERAL INFORMATION:
APPLICANT: Seed, Brian et al.
TITLE OF INVENTION: Redirection of Cellular Immunity by
Receptor Chimeras

NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
COMPUTER: IBM PS/2 Model 50Z or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: Wordperfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/243,008
FILING DATE: 02-Feb-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/394,176
FILING DATE: SEPTEMBER 11,1995
APPLICATION NUMBER: 08/203,866
FILING DATE: February 28, 1994
APPLICATION NUMBER: 07/847,566
FILING DATE: March 6, 1992
APPLICATION NUMBER: 07/665,961
FILING DATE: March 7, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Karen F. Lech, Ph.D.
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/270001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-243-008-6
Alignment Scores:
Pred. No.: 299 Length: 532
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 11 Gaps: 0
US-10-029-345A-108_COPY_532_2532 (1-2001) x US-09-243-008-6 (1-532)
QY 1819 GACGAGCTGACTGCGGCGGAGC 1842
DB 78 AspArgAlaAspSerArgArgSer 85
RESULT 38
US-10-046-232-22
Sequence 22, Application US/10046232
Publication No. US20030119099A1
GENERAL INFORMATION:
APPLICANT: Helmut SCHWAB
APPLICANT: Anton GLEDER
APPLICANT: Christoph KRATKY
APPLICANT: Ingrid DREVENY
APPLICANT: Peter POCHLAUER
APPLICANT: Wolfgang SKRANC
APPLICANT: Herbert MAYRHOFER
APPLICANT: Irma WIRTH
APPLICANT: Rudolf NEUHOFER
APPLICANT: Rodolfo BONA
TITLE OF INVENTION: New genes containing a DNA sequence coding for a hydroxynitrile 1
recombinant proteins derived therefrom and having hydroxynitrile

TITLE OF INVENTION: thereof
FILE REFERENCE: 2001-1982A/LC/01553
CURRENT APPLICATION NUMBER: US/10/046,232
CURRENT FILING DATE: 2002-10-31
PRIOR APPLICATION NUMBER: A60/2001
PRIOR FILING DATE: 2001-01-16
PRIOR APPLICATION NUMBER: A523/2001
PRIOR FILING DATE: 2001-04-03
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 22
LENGTH: 534
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of the artificial sequence: Hybrid protein PamHNL5xGC
US-10-046-232-22

Alignment Scores:
Pred. No.: 299 Length: 534
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 14 Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x US-10-046-232-22 (1-534)

QY 386 AGTTCTGCTGTTCTTCCTGACC 409

Db 7 SerSerLeuValValSerLeuVal 14

RESULT 39

US-10-156-761-11387
Sequence 11387, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMIURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 11387
LENGTH: 544
TYPE: PRT
ORGANISM: Streptomyces avermilti1115
US-10-156-761-11387

Alignment Scores:
Pred. No.: 298 Length: 544
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 14 Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x US-10-156-761-11387 (1-544)

QY 1648 ACCTTACCCCTTCCTGACACAC 1671

Db 13 ThrSerThrProSerLeuThrSer 20

RESULT 40

US-10-628-088-415
Sequence 415, Application US/10628088
Publication No. US20040096451A1
GENERAL INFORMATION:
APPLICANT: Young, James
APPLICANT: Kiener, Peter
APPLICANT: Osterhaus, Albertus
APPLICANT: Fouchier, Ronaldus
TITLE OF INVENTION: METHODS OF TREATING AND PREVENTING
TITLE OF INVENTION: RSV, HMPV, AND PIV USING ANTI-RSV,
TITLE OF INVENTION: ANTI-HMPV, AND ANTI-PIV ANTIBODIES
FILE REFERENCE: 10271-072-999
CURRENT APPLICATION NUMBER: US/10/628,088
CURRENT FILING DATE: 2003-07-25
PRIOR APPLICATION NUMBER: 60/398,475
PRIOR FILING DATE: 2002-07-25
NUMBER OF SEQ ID NOS: 437
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 415
LENGTH: 555
TYPE: PRT
ORGANISM: Human parainfluenza virus 1 strain Washington/1964
FEATURE:
OTHER INFORMATION: F glycoprotein of Human parainfluenza 1 virus
US-10-628-088-415

Alignment Scores:
Pred. No.: 298 Length: 555
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0
DB: 16 Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x US-10-628-088-415 (1-555)

QY 1193 AGCTTATGCTGCTTCCTCCAGCCTG 1170

Db 13 SerLeuLeuSerSerSerLeu 20

RESULT 41

US-09-939-537-4
Sequence 4, Application US/09939537
Publication No. US20030138410A1
GENERAL INFORMATION:
APPLICANT: Seed, Brian
APPLICANT: Banapour, Babak
APPLICANT: Romeo, Charles
TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
CELLS BY CHIMERIC CD4 RECEPTOR-BEARING CELLS
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Clark & Elbing LLP
STREET: 176 Federal Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/939,537
FILING DATE: 24-Aug-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/284,391
FILING DATE: 02-AUG-1994
APPLICATION NUMBER: 08/195,395

FILING DATE: 14-FEB-1994
APPLICATION NUMBER: 07/847,566
FILING DATE: 06-MAR-1992
APPLICATION NUMBER: 07/665,961
FILING DATE: 07-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Elding, Karen L.
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/247001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045
TELEX: <unknown>
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 575 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-939-537-4

Alignment Scores:
Pred. No.: 296 Length: 575
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 10 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-09-939-537-4 (1-575)

QY 1819 GACAGAGCTGACTCGCGCGGAGC 1842
DB 78 AsparGalaAspSerArgSer 85

RESULT 42

US-09-243-008-4
Sequence 4, Application US/09243008
Publication No. US20040005334A1
GENERAL INFORMATION:
APPLICANT: Seed, Brian et al.
TITLE OF INVENTION: Redirection of Cellular Immunity by
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 50Z or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: Wordperfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/243,008
FILING DATE: 02-Feb-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/394,176
FILING DATE: SEPTEMBER 11, 1995
APPLICATION NUMBER: 08/203,866
FILING DATE: February 28, 1994
APPLICATION NUMBER: 07/847,566
FILING DATE: March 6, 1992
APPLICATION NUMBER: 07/665,961
FILING DATE: March 7, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Karen F. Lech, Ph.D
REGISTRATION NUMBER: 35,238

REFERENCE/DOCKET NUMBER: 00786/270001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 575 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-243-008-4

Alignment Scores:
Pred. No.: 296 Length: 575
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 11 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-09-243-008-4 (1-575)

QY 1819 GACAGAGCTGACTCGCGCGGAGC 1842
DB 78 AsparGalaAspSerArgSer 85

RESULT 43

US-09-934-060A-13
Sequence 13, Application US/09934060A
Patent No. US20020155121A1
GENERAL INFORMATION:
APPLICANT: Devico, Anthony L.
APPLICANT: Fouts, Timothy R.
APPLICANT: Tuskan, Robert G.
TITLE OF INVENTION: VIRUS COAT PROTEIN/RECEPTOR CHIMERAS AND METHODS OF USE
FILE REFERENCE: 4115-144 CIP
CURRENT APPLICATION NUMBER: US/09/934,060A
CURRENT FILING DATE: 2001-08-21
PRIOR APPLICATION NUMBER: US 09/684,026
PRIOR FILING DATE: 2000-10-06
PRIOR APPLICATION NUMBER: US 60/158,321
PRIOR FILING DATE: 1999-10-08
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn version 3.1
SEQ ID NO 13
LENGTH: 590
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (586)..(586)
OTHER INFORMATION: Xaa can be any amino acid
NAME/KEY: MISC FEATURE
LOCATION: (589)..(589)
OTHER INFORMATION: Xaa can be any amino acid
US-09-934-060A-13

Alignment Scores:
Pred. No.: 296 Length: 590
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 9 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-09-934-060A-13 (1-590)

QY 1819 GACAGAGCTGACTCGCGCGGAGC 1842
DB 450 AsparGalaAspSerArgSer 457

```
RESULT 44
US-10-062-937B-5
; Sequence 5, Application US/10062937B
; Publication No. US20030022195A1
; GENERAL INFORMATION:
; APPLICANT: Curtis, Roy A.J.
; TITLE OF INVENTION: 59914 and 59921, CHOLINE TRANSPORTERS
; TITLE OF INVENTION: AND USES THEREFOR
; FILE REFERENCE: MP12001-005PIRM
; CURRENT APPLICATION NUMBER: US/10/062,937B
; CURRENT FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: 60/267,076
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 653
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-062-937B-5

Alignment Scores:
Pred. No.:      292      Length:      653
Score:          8.00     Matches:      8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:    1.24%   Indels:      0
DB:             14      Gaps:       0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-10-062-937B-5 (1-653)
QY      1493 TGGGCGGTGCGCACTGCTGCTGTT 1470
DB      564 TTPAlaValProLeuLeuVal 571

RESULT 45
US-10-391-399-73
; Sequence 73, Application US/10391399
; Publication No. US20030219806A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Glucksmann, Maria Alexandra
; APPLICANT: Curtis, Roy A. J.
; APPLICANT: Lora, Jose M.
; APPLICANT: Galvin, Katherine M.
; APPLICANT: Silos-Santiago, Immaculada
; TITLE OF INVENTION: NOVEL 18607, 15603, 69318, 12303, 48000,
; TITLE OF INVENTION: 52920, 5433, 38554, 57301, 58324, 55063, 52991, 59914, 59921
; TITLE OF INVENTION: AND 33751 MOLECULES AND USES THEREFOR
; FILE REFERENCE: MP103-020OMIM
; CURRENT APPLICATION NUMBER: US/10/391,399
; CURRENT FILING DATE: 2003-03-18
; PRIOR APPLICATION NUMBER: US 09/789,481
; PRIOR FILING DATE: 2001-02-20
; PRIOR APPLICATION NUMBER: US 09/634,669
; PRIOR FILING DATE: 2000-08-08
; PRIOR APPLICATION NUMBER: US 09/583,373
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: US 09/510,706
; PRIOR FILING DATE: 2000-02-22
; PRIOR APPLICATION NUMBER: US 10/309,804
; PRIOR FILING DATE: 2002-12-04
; PRIOR APPLICATION NUMBER: US 60/336,936
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: US 10/094,214
; PRIOR FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: US60/275,078
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 09/828,035
; PRIOR FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: US 60/195,734
; PRIOR FILING DATE: 2000-04-07
```

```
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 127
; SOFTWARE: FaastSeq for Windows Version 4.0
; SEQ ID NO 73
; LENGTH: 653
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-391-399-73

Alignment Scores:
Pred. No.:      292      Length:      653
Score:          8.00     Matches:      8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:    1.24%   Indels:      0
DB:             15      Gaps:       0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-10-391-399-73 (1-653)
QY      1493 TGGGCGGTGCGCACTGCTGCTGTT 1470
DB      564 TTPAlaValProLeuLeuVal 571

RESULT 46
US-09-833-745-39
; Sequence 39, Application US/09833745
; Patent No. US20020052038A1
; GENERAL INFORMATION:
; APPLICANT: ROBERTS, JOSEPH
; APPLICANT: SETHURAMAN, NATARAJAN
; APPLICANT: MACALLISTER, THOMAS
; TITLE OF INVENTION: CLONING, OVEREXPRESSION AND THERAPEUTIC USE OF
; TITLE OF INVENTION: BIOACTIVE HISTIDINE AMMONIA LYASE
; FILE REFERENCE: 078728/0106
; CURRENT APPLICATION NUMBER: US/09/833,745
; CURRENT FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/197,770
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 66
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 39
; LENGTH: 677
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-09-833-745-39

Alignment Scores:
Pred. No.:      291      Length:      677
Score:          8.00     Matches:      8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match:    1.20%   Indels:      0
DB:             9       Gaps:       0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-09-833-745-39 (1-677)
QY      137 AAGCATTATATCAACTGCTCCA 160
DB      584 LysProLeuIleSerThrAlaPro 591

RESULT 47
US-10-369-493-6901
; Sequence 6901, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
```

```
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 6901
; LENGTH: 711
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-6901

Alignment Scores:
Pred. No.: 291      Length: 677
Score: 8.00      Matches: 8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 1.20%      Indels: 0
DB: 15      Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-10-369-493-6901 (1-677)
QY 137 AAGCATTATATCACTGCTCCA 160
Db 584 LysProleulinSerThrAlaPro 591

RESULT 48
US-09-792-630-33
; Sequence 33, Application US/09792630
; Patent No. US20020168640A1
; GENERAL INFORMATION:
; APPLICANT: Li, Min
; TITLE OF INVENTION: BIOCHIPS COMPRISING NUCLEIC ACID/PROTEIN CONJUGATES
; FILE REFERENCE: A-70295/RFT/RMS/RMK
; CURRENT APPLICATION NUMBER: US/09/792,630
; CURRENT FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
; LENGTH: 711
; TYPE: PRT
; ORGANISM: chipmunk parvovirus
US-09-792-630-33

Alignment Scores:
Pred. No.: 290      Length: 711
Score: 8.00      Matches: 8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 1.24%      Indels: 0
DB: 9      Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-09-792-630-33 (1-711)
QY 1647 CTGGGGGGCCAGATATCCGAGTG 1624
Db 404 LeuGlyGlyGlnAspIleArgVal 411

RESULT 49
US-09-953-351-33
; Sequence 33, Application US/09953351
; Publication No. US20030036643A1
; GENERAL INFORMATION:
; APPLICANT: Li, Min
; APPLICANT: Melander, Christian
; APPLICANT: Liu, Hong-Xiang
; APPLICANT: Jin, Cheng He
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE CONSTRUCTION AND USE OF FUSION I
; FILE REFERENCE: A-70814/RFT/RMS/RMK
; CURRENT APPLICATION NUMBER: US/09/953,351
; CURRENT FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: US 60/232,960
; PRIOR FILING DATE: 2000-09-14
```

```
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
; LENGTH: 711
; TYPE: PRT
; ORGANISM: chipmunk parvovirus
US-09-953-351-33

Alignment Scores:
Pred. No.: 290      Length: 711
Score: 8.00      Matches: 8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 1.24%      Indels: 0
DB: 10      Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-09-953-351-33 (1-711)
QY 1647 CTGGGGGGCCAGATATCCGAGTG 1624
Db 404 LeuGlyGlyGlnAspIleArgVal 411

RESULT 50
US-10-080-376-33
; Sequence 33, Application US/10080376
; Publication No. US20020172968A1
; GENERAL INFORMATION:
; APPLICANT: Li, Min
; APPLICANT: Dahiyat, Basail I.
; TITLE OF INVENTION: BIOCHIPS COMPRISING NUCLEIC ACID/PROTEIN CONJUGATES
; FILE REFERENCE: A-70295-2/RFT/RMS/RMK
; CURRENT APPLICATION NUMBER: US/10/080,376
; CURRENT FILING DATE: 2000-02-19
; PRIOR APPLICATION NUMBER: US 09/792,630
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
; LENGTH: 711
; TYPE: PRT
; ORGANISM: chipmunk parvovirus
US-10-080-376-33

Alignment Scores:
Pred. No.: 290      Length: 711
Score: 8.00      Matches: 8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 1.24%      Indels: 0
DB: 13      Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-10-080-376-33 (1-711)
QY 1647 CTGGGGGGCCAGATATCCGAGTG 1624
Db 404 LeuGlyGlyGlnAspIleArgVal 411

Search completed: June 21, 2004, 13:19:42
Job time : 166.5 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 21, 2004, 13:05:47 ; Search time 24.5 Seconds
(without alignments)
8432.944 Million cell updates/sec

Title: US-10-029-345A-108_COPY_532_2532

Perfect score: 667
Sequence: 1 gctgcacatgcagccatgcagat.....tggaaatcatctgagctcc 2001

Scoring table:
OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues

Word size: 1

Total number of hits satisfying chosen parameters: 56994

Minimum DB seq length: 473

Maximum DB seq length: 200000000

Post-processing: Listing first 75 summaries

Command line parameters:
-MODE=frame_n2p.model -DEV=xlp
-Q/cg2_1/USPTO.spool.p/US10029345/funat.21062004.124616.5729/app.query.fasta_1.2183
-DB=Issued Patents AA -QFMT=fasta -SUFFIX=oligo.ra -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=75 -DOALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=50 -MODE=LOCAL
-OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=473 -MAXLEN=200000000
-USER=US10029345 -CCN 1 1 27 -runat.21062004.124616.5729 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -Fgapop=6
-FAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued Patents AA:*

1: /cg2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cg2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cg2_6/prodata/2/1aa/5A_COMB.pep:*
4: /cg2_6/prodata/2/1aa/5B_COMB.pep:*
5: /cg2_6/prodata/2/1aa/PCTUS_COMB.pep:*
6: /cg2_6/prodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	472	70.8	665	4 US-09-816-494-2	Sequence 2, Appl1
2	10	1.6	503	4 US-09-134-000C-5978	Sequence 5978, Ap
3	3	1.4	737	4 US-08-989-299-6	Sequence 6, Appl1
4	4	1.4	737	4 US-09-407-427-6	Sequence 11, Appl1
5	5	1.2	503	4 US-08-740-223A-11	Sequence 11, Appl1
6	6	1.2	503	4 US-09-709-188-11	Sequence 8, Appl1
7	7	1.2	509	2 US-08-665-326-8	Sequence 10, Appl1
8	8	1.2	509	3 US-08-740-223A-10	Sequence 2, Appl1
9	9	1.2	509	4 US-09-202-491-2	Sequence 10, Appl1
10	10	1.2	509	4 US-09-202-491-3	Sequence 3, Appl1
11	11	1.2	509	4 US-09-709-188-10	Sequence 10, Appl1
12	12	1.2	511	4 US-09-252-991A-18691	Sequence 18691, A

13	8	1.2	530	3 US-08-477-460B-4	Sequence 4, Appl1
14	8	1.2	530	3 US-08-379-516-4	Sequence 4, Appl1
15	8	1.2	530	3 US-09-329-916-4	Sequence 4, Appl1
16	8	1.2	530	3 US-08-483-372A-4	Sequence 4, Appl1
17	8	1.2	530	4 US-09-409-006A-4	Sequence 4, Appl1
18	8	1.2	530	4 US-08-484-681-4	Sequence 4, Appl1
19	8	1.2	530	5 PCT-US93-07422-4	Sequence 4, Appl1
20	8	1.2	532	2 US-08-417-495-6	Sequence 6, Appl1
21	8	1.2	532	2 US-08-284-391B-6	Sequence 6, Appl1
22	8	1.2	532	2 US-09-218-950-6	Sequence 6, Appl1
23	8	1.2	532	5 PCT-US92-01785-6	Sequence 6, Appl1
24	8	1.2	532	5 PCT-US95-00454-6	Sequence 6, Appl1
25	8	1.2	575	2 US-08-417-495-4	Sequence 4, Appl1
26	8	1.2	575	2 US-08-284-391B-4	Sequence 4, Appl1
27	8	1.2	575	2 US-09-218-950-6	Sequence 4, Appl1
28	8	1.2	575	5 PCT-US92-01785-4	Sequence 4, Appl1
29	8	1.2	575	5 PCT-US95-00454-4	Sequence 4, Appl1
30	8	1.2	592	4 US-09-252-991A-18124	Sequence 18124, A
31	8	1.2	604	2 US-08-746-283-12	Sequence 32, Appl1
32	8	1.2	604	2 US-08-746-257A-30	Sequence 30, Appl1
33	8	1.2	605	1 US-08-333-802-2	Sequence 2, Appl1
34	8	1.2	605	4 US-09-347-878-46	Sequence 46, Appl1
35	8	1.2	607	4 US-09-252-991A-26137	Sequence 26137, A
36	8	1.2	611	4 US-08-746-283-31	Sequence 28230, A
37	8	1.2	612	2 US-08-746-257A-29	Sequence 29, Appl1
38	8	1.2	612	2 US-08-746-283-1	Sequence 1, Appl1
39	8	1.2	613	2 US-08-746-283-3	Sequence 1, Appl1
40	8	1.2	613	2 US-08-746-283-5	Sequence 5, Appl1
41	8	1.2	613	2 US-08-746-257A-1	Sequence 1, Appl1
42	8	1.2	613	4 US-09-347-878-48	Sequence 48, Appl1
43	8	1.2	613	4 US-08-472-888A-6	Sequence 6, Appl1
44	8	1.2	683	4 US-09-591-095-6	Sequence 6, Appl1
45	8	1.2	903	1 US-08-021-601-12	Sequence 12, Appl1
46	8	1.2	903	1 US-08-082-849B-12	Sequence 12, Appl1
47	8	1.2	903	5 PCT-US94-01642-12	Sequence 12, Appl1
48	8	1.2	935	4 US-09-252-991A-33037	Sequence 33037, A
49	8	1.2	953	4 US-08-252-991A-27230	Sequence 27230, A
50	8	1.2	2500	2 US-08-801-263A-2	Sequence 2, Appl1
51	8	1.2	2500	3 US-09-102-248-2	Sequence 2, Appl1
52	8	1.2	2500	4 US-09-367-764-2	Sequence 9, Appl1
53	8	1.2	2512	2 US-08-801-263A-9	Sequence 9, Appl1
54	8	1.2	2512	3 US-09-102-248-9	Sequence 9, Appl1
55	8	1.2	2512	4 US-09-367-764-9	Sequence 9, Appl1
56	8	1.2	2517	2 US-08-801-263A-5	Sequence 5, Appl1
57	8	1.2	2517	3 US-09-102-248-5	Sequence 5, Appl1
58	8	1.2	2517	4 US-09-367-764-5	Sequence 5, Appl1
59	8	1.2	2517	1 US-07-741-940-7	Sequence 7, Appl1
60	8	1.2	2542	1 US-08-452-654A-7	Sequence 7, Appl1
61	8	1.2	2542	1 US-08-452-654B-7	Sequence 7, Appl1
62	8	1.2	2542	1 US-08-452-654-7	Sequence 7, Appl1
63	8	1.2	2542	1 US-08-449-731-7	Sequence 7, Appl1
64	8	1.2	2543	4 US-07-741-940-2	Sequence 2, Appl1
65	8	1.2	2543	1 US-08-289-548A-2	Sequence 2, Appl1
66	8	1.2	2543	1 US-08-452-654A-2	Sequence 2, Appl1
67	8	1.2	2543	1 US-08-452-655B-2	Sequence 2, Appl1
68	8	1.2	2543	1 US-08-452-655B-7	Sequence 7, Appl1
69	8	1.2	2543	2 US-08-370-235A-2	Sequence 2, Appl1
70	8	1.2	2543	3 US-08-450-582-2	Sequence 2, Appl1
71	8	1.2	2543	3 US-08-450-582-7	Sequence 7, Appl1
72	8	1.2	2543	4 US-08-449-731-2	Sequence 2, Appl1
73	8	1.2	2573	2 US-08-821-355A-7	Sequence 7, Appl1
74	8	1.2	2573	2 US-09-003-687A-7	Sequence 7, Appl1
75	8	1.2	2573	3 US-09-136-605-7	Sequence 7, Appl1

ALIGNMENTS

RESULT 1
US-09-816-494-2
; Sequence 2, Application US/09816494
; Patent No. 6664089
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel A.

TITLE OF INVENTION: 38692 AND 21117, NOVEL DUAL SPECIFICITY
FILE REFERENCE: 10448-030002
CURRENT APPLICATION NUMBER: US/09/816,494
PRIOR APPLICATION NUMBER: 2001-03-23
PRIOR FILING DATE: 2000-03-24
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 665
TYPE: PRT
ORGANISM: Homo sapiens
US-09-816-494-2

Alignment Scores:
Pred. No.: 0 Length: 665
Score: 472.00 Matches: 472
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 70.76% Indels: 0
Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-09-816-494-2 (1-665)

QY 586 ACCTGTCGCAAGCTGATTTATCCCGAGTCTCATTTCTGCGTGTGAGTAAGAC 645
Db 194 ThrcysproulspromaphhellleProgluserishshelenuhrgvalProvalaamsp 213
QY 646 AGCTTTGTGAGAAATTTTCCCGTGTGAGCAAAATCAGTAAATTTCAATTTGAGAAAGCA 705
Db 214 Serphecylulyslleleuprotiprleuaspulyservalaspheilleglulysala 233
QY 706 AAAGCCTCAATGATGTGTCTAGTCACTGTTTGGAGTCCCGCTCCGCGACAC 765
Db 234 Lyelaserhenglycysvalleuvalhiscyleuvaladilylleserargseralatr 253
QY 766 ATGCTTATGCTTCATCATCAAGAGAGATGACATGCTTTAGATGAAGCTTACAGATT 825
Db 254 llelallelalellyrillemetlysarvmelaspmetserleuaspulualatyrargphe 273
QY 826 GTGAAAGAAAAGAAAGCTATATATCTCCAACTTGAATTTTGTGGCCCACTCTGGAC 885
Db 274 VallysglulysargpromtrilleseProaamphenelenuylglinleuenuasp 293
QY 886 TATGAGAAAGATTAAGAACAGACTGAGCATCAAGCCCAAGCAACTCAAGCTG 945
Db 294 TyrglulyselylellyshenglinthrgylalaserdlyProlyserllyseulysueu 313
QY 946 CTGCACCTGAGAAAGCAATGAACTGTTCCTGCTGTCTCAAGGGGTGACAGAAAAC 1005
Db 314 LeuhisLeuenglulysProaangluprovalProalavalserglulylglinlysser 333
QY 1006 GAGACGCCCTCAGTCACTGTGCGCACTGCGCACTCGAGAGAGAGAGAGAAAG 1065
Db 334 GluthrProleuuserProcyosalaaapseralathmserglualaladilylindrg 353
QY 1066 CCCGTCATCCCGCCAGCGTGCACAGCTGCCCAAGCTGACCGCTGCTTGAAGAGAC 1125
Db 354 ProvalhlsproalaservalProservalProservalGlnProserleuenglunsp 373
QY 1126 AGCCGCTGTAAGAGCGCTGAGTGGGTGACCTGTCCGAGAGAGAGCTGAGAGACGC 1185
Db 374 SerproleuvalGlnhlaeuserdlyleuhisleuseralaaapargleuGlunspser 393
QY 1186 AATAAGCTCAAGCTTCTTCTCTGATATCAATCAATGATTTCAATGAGCAGCATG 1245
Db 394 AenlyslleuylargserPheuserleuunspilleyservalSerlyrseralSermet 413
QY 1246 GAGAGATCTTACATGAGCTTCTCTCATCAAGAAAGCTTTGAATACTACAAACTTTC 1305
Db 414 AlaalaSerleuuhsglyPheuserserserglunspalatalengluluyrlylyseProser 433

QY 1306 ACTACTCTGATGGAGCAACAAGCTATCCAGTTCTCCCTGTTCAAGAACTATGAGAC 1365
Db 434 ThrtThreleuaspolythraenlyseuysglinpneserProvalGlnleuuserglu 453
QY 1366 CAGACTCCCGAAAACAGGTCTCTATAGAGAGAGAGAGAGATCCCAAGAAAGCTGAGACC 1425
Db 454 GluthrProgluthserProaspulysglulualaserilleProlyseulysleuInthr 473
QY 1426 GCCAGGCTTCAAGACAGCAGAGCAAGAGATTTGATTCGATGAGAGAGAGAGAGAGAG 1485
Db 474 AlaargProserAspsercInserlysargleuuhiservalAargThrserseuergly 493
QY 1486 ACCGCCCAAGGTCCTTTTATCTTCACCTGATGAGAGTGGAGGCTGAGAGACATTAAC 1545
Db 494 ThralaGlnargserleuuserProleuhisargsergllysevalGluaspamlyr 513
QY 1546 CACACAGCTTCTTTTGGGCTTTCCACAGCCAGAGACACCTTACCAAGTCTGCTGCG 1605
Db 514 HistrserPheleuPhegllyseuuserThrserGlnGlnhlsleuthrlysservaladly 533
QY 1606 CTGGGCTTAAAGGCTGACCTCGATATCTTGGCCCCCAGACCTTACCCCTTCCCTG 1665
Db 534 leuglyleuylsglyTtrpHisseraspilleuualaroglinthserThrProserleu 553
QY 1666 ACCAGACCTGGTATTTTGGCCAGAGATCTTCAACTTTACTGTGCTCAGCCATCTAC 1725
Db 554 ThrserserlyrlyrPhelalathrGluserserhlspheryrseralaseralalelyr 573
QY 1726 GGAGGACGTGCGATTACTGTGCTTACAGCTGCGAGCCAGCTGCCCATTTGGGAGACCA 1785
Db 574 GlylseralaserlyrseralalyrserCysserleuenuprohmCysgllyaspGln 593
QY 1786 GTCTATTGTTCGCGAGGCGGCGAGAGGCAAGTCAAGAGCTGCTCGCGGAGCTGG 1845
Db 594 ValtyrserValargargargGlnlyProseraspbrglaaapserargargserTtrp 613
QY 1846 CATGAAAGAGACCCCTTTGAAAAGCATTTAAAGCGAAGCTGCCAAATGAAATTTGGA 1905
Db 614 HlsgluluserProphleglulysGlnphelysarargrgerCysGlnmetGluPhegly 633
QY 1906 GAGAGCATCATGTGAGAGAACAGGTCAAGGAGAGAGAGCTGGGAGAAAGTGGAGCTGACT 1965
Db 634 GluserillemetserGluuamargserargGlnGlnleuengllysevalGlyserGlnser 653
QY 1966 AGCTTTCCGGCAGCATGAAATCATTTGAGGCTTCC 2001
Db 654 SerpheserGlysermetGluilleleGluvaliser 665

RESULT 2

US-09-134-000C-5978
Sequence 5978, Application US/09134000C
Patent No. 6617156
GENERAL INFORMATION:
APPLICANT: Lynn Doucelte-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
FILE REFERENCE: 032796-032
CURRENT APPLICATION NUMBER: US/09/134,000C
PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/055,778
NUMBER OF SEQ ID NOS: 6812
SOFTWARE: PatentIn version 3.1
SEQ ID NO 5978
LENGTH: 503
TYPE: PRT
ORGANISM: Enterococcus faecalis
US-09-134-000C-5978

Alignment Scores:
Pred. No.: 0.515 Length: 503
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.55% Indels: 0
DB: 4 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-09-134-000C-5978 (1-503)

QY 103 GCCGCTATCATTTAGCAGCATTTCCTCCG 74
DB 347 AAGGlyTTCGlnLeuAlaIleuPhePro 356

RESULT 3
US-08-989-299-6
; Sequence 6, Application US/08989299
; Patent No. 6194556
; GENERAL INFORMATION:
; APPLICANT: Acton, Susan L.
; APPLICANT: Robinson, Keith E.
; TITLE OF INVENTION: ANGIOTENSIN CONVERTING ENZYME HOMOLOG
; TITLE OF INVENTION: ANGIOTENSIN CONVERTING ENZYME HOMOLOG
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/989,299
; FILING DATE: 11-DEC-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold E. Beth
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: MIA-025.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 737 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-989-299-6

Alignment Scores:
Pred. No.: 4.6 Length: 737
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.40% Indels: 0
DB: 3 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-08-989-299-6 (1-737)

QY 1334 CATAGCTGTGTCCTCCATCCAGAGTA 1308
DB 22 HisSerLeuValProSerArgVal 30

RESULT 4
US-09-407-427-6
; Sequence 6, Application US/09407427
; Patent No. 6610497
; GENERAL INFORMATION:
; APPLICANT: Acton, Susan L.
; APPLICANT: Robinson, Keith E.

; TITLE OF INVENTION: ANGIOTENSIN CONVERTING ENZYME HOMOLOG AND THERAPEUTIC
; TITLE OF INVENTION: AND DIAGNOSTIC USES THEREFOR
; FILE REFERENCE: MNT-132CP2
; CURRENT APPLICATION NUMBER: US/09/407,427
; CURRENT FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: 09/163,648
; PRIOR FILING DATE: 1998-09-30
; PRIOR APPLICATION NUMBER: 08/989,299
; PRIOR FILING DATE: 1997-12-11
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 6
; LENGTH: 737
; TYPE: PRT
; ORGANISM: Oryctolagus cuniculus
; US-09-407-427-6

Alignment Scores:
Pred. No.: 4.6 Length: 737
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.40% Indels: 0
DB: 4 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-09-407-427-6 (1-737)

QY 1334 CATAGCTGTGTCCTCCATCCAGAGTA 1308
DB 22 HisSerLeuValProSerArgVal 30

RESULT 5
US-08-740-223A-11
; Sequence 11, Application US/08740223A
; Patent No. 6265564
; GENERAL INFORMATION:
; APPLICANT: Davis, et al.
; TITLE OF INVENTION: Expressed ligand - Vascular
; TITLE OF INVENTION: Intracellular Signalling Molecule
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Regeneron Pharmaceuticals, Inc.
; STREET: 777 Old Saw Mill Road
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/740,223A
; FILING DATE: 25-OCT-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: USSN 60/022/999
; FILING DATE: 02-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Cobert, Robert J
; REGISTRATION NUMBER: 36,108
; REFERENCE/DOCKET NUMBER: REG 333
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 914-345-7400
; TELEFAX: 914-345-7721
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 503 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein

```
/
/ FEATURE:
/ NAME/KEY: mtl3
/ LOCATION: 1..503
/ OTHER INFORMATION: mouse TIE ligand-3
US-08-740-223A-11

Alignment Scores:
Pred. No.: 44.5 Length: 503
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 3 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-08-740-223A-11 (1-503)
QY 1083 CGTGGCCAGCGTGGCCAGCGTGCA 1106
DB 80 ArgAlaGlnArgAlaGlnArgAla 87

RESULT 6
US-09-709-188-11
/ Sequence 11, Application US/09709188
/ Patent No. 6441137
/ GENERAL INFORMATION:
/ APPLICANT: Davis et al.
/ TITLE OF INVENTION: Expressed Ligand - Vascular InterCellular Signaling Molecule
/ FILE REFERENCE: REG 333-2
/ CURRENT APPLICATION NUMBER: US/09/709,188
/ PRIOR FILING DATE: 2000-11-09
/ PRIOR APPLICATION NUMBER: 08/740,223
/ NUMBER OF SEQ ID NOS: 30
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 11
/ LENGTH: 503
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-709-188-11

Alignment Scores:
Pred. No.: 44.5 Length: 503
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 4 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-09-709-188-11 (1-503)
QY 1083 CGTGGCCAGCGTGGCCAGCGTGCA 1106
DB 80 ArgAlaGlnArgAlaGlnArgAla 87

RESULT 7
US-08-665-926-8
/ Sequence 8, Application US/08665926
/ Patent No. 5851797
/ GENERAL INFORMATION:
/ APPLICANT: Valenzuela et al.
/ TITLE OF INVENTION: TIE Ligand-3, METHODS OF MAKING AND USES
/ NUMBER OF SEQUENCES: 8
/ CORRESPONDENCE ADDRESS:
/ ADDRESSER: Regeneron Pharmaceuticals, Inc.
/ STREET: 777 Old Saw Mill River Road
/ CITY: Tarrytown
/ STATE: New York
/ COUNTRY: U.S.A.
/ ZIP: 10591-6707
/ TELEPHONE: 914-345-7400
/ TELEFAX: 914-345-7421
/ INFORMATION FOR SEQ ID NO: 10:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ LENGTH: 509 amino acids
```

```
/
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/665,926
/ FILING DATE: 19-JUN-1996
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Robert J. Covert
/ REGISTRATION NUMBER: 36,108
/ REFERENCE/DOCKET NUMBER: REG 330-H
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (914) 345-7400
/ TELEFAX: (914) 345-2113
/ INFORMATION FOR SEQ ID NO: 8:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 509 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
US-08-665-926-8

Alignment Scores:
Pred. No.: 44.4 Length: 509
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 2 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-08-665-926-8 (1-509)
QY 1083 CGTGGCCAGCGTGGCCAGCGTGCA 1106
DB 86 ArgAlaGlnArgAlaGlnArgAla 93

RESULT 8
US-08-740-223A-10
/ Sequence 10, Application US/08740223A
/ Patent No. 6265564
/ GENERAL INFORMATION:
/ APPLICANT: Davis, et al.
/ TITLE OF INVENTION: Expressed Ligand - Vascular
/ TITLE OF INVENTION: InterCellular Signalling Molecule
/ NUMBER OF SEQUENCES: 28
/ CORRESPONDENCE ADDRESS:
/ ADDRESSER: Regeneron Pharmaceuticals, Inc.
/ STREET: 777 Old Saw Mill Road
/ CITY: Tarrytown
/ STATE: NY
/ COUNTRY: USA
/ ZIP: 10591
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/ SOFTWARE: FastSeq Version 2.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/740,223A
/ FILING DATE: 25-OCT-1996
/ CLASSIFICATION: 536
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: USSN 60/022/999
/ FILING DATE: 02-AUG-1996
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Covert, Robert J
/ REGISTRATION NUMBER: 36,108
/ REFERENCE/DOCKET NUMBER: REG 333
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 914-345-7400
/ TELEFAX: 914-345-7421
/ INFORMATION FOR SEQ ID NO: 10:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ LENGTH: 509 amino acids
```



```
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; FEATURE:
; NAME/KEY: TIE ligand-3
; LOCATION: 1...509
; OTHER INFORMATION:
US-08-740-223a-10

Alignment Scores:
Pred. No.: 44.4
Score: 8.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 1.20%
DB: 3
Length: 509
Matches: 8
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x US-08-740-223a-10 (1-509)
QY 1083 CGTCCCGAGCGTCCCGCGCTGCA 1106
DB 86 ArgAlaGlnArgAlaGlnArgAla 93

RESULT 9
US-09-202-491-2
; Sequence 2, Application US/09202491
; Patent No. 643267
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al.
; TITLE OF INVENTION: NOVEL LIGANDS, METHODS OF MAKING AND USES THEREOF
; FILE REFERENCE: REG330-K
; CURRENT APPLICATION NUMBER: US/09/202,491
; EARLIER FILING DATE: 1998-11-16
; EARLIER APPLICATION NUMBER: PCT/US97/10728
; EARLIER FILING DATE: 1997-06-19
; EARLIER APPLICATION NUMBER: 60/022,999
; EARLIER FILING DATE: 1996-08-02
; EARLIER APPLICATION NUMBER: 60/021,087
; EARLIER FILING DATE: 1996-07-02
; EARLIER APPLICATION NUMBER: 08/665,926
; EARLIER FILING DATE: 1996-06-19
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-202-491-2

Alignment Scores:
Pred. No.: 44.4
Score: 8.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 1.20%
DB: 4
Length: 509
Matches: 8
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x US-09-202-491-2 (1-509)
QY 1083 CGTCCCGAGCGTCCCGCGCTGCA 1106
DB 86 ArgAlaGlnArgAlaGlnArgAla 93

RESULT 10
US-09-202-491-3
; Sequence 3, Application US/09202491
; Patent No. 643267
; GENERAL INFORMATION:
; APPLICANT: Valenzuela et al.
; TITLE OF INVENTION: NOVEL LIGANDS, METHODS OF MAKING AND USES THEREOF
; FILE REFERENCE: REG330-K
```

```
; CURRENT APPLICATION NUMBER: US/09/202,491
; EARLIER FILING DATE: 1998-11-16
; EARLIER APPLICATION NUMBER: PCT/US97/10728
; EARLIER FILING DATE: 1997-06-19
; EARLIER APPLICATION NUMBER: 60/022,999
; EARLIER FILING DATE: 1996-08-02
; EARLIER APPLICATION NUMBER: 60/021,087
; EARLIER FILING DATE: 1996-07-02
; EARLIER APPLICATION NUMBER: 08/665,926
; EARLIER FILING DATE: 1996-06-19
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-202-491-3

Alignment Scores:
Pred. No.: 44.4
Score: 8.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 1.20%
DB: 4
Length: 509
Matches: 8
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x US-09-202-491-3 (1-509)
QY 1083 CGTCCCGAGCGTCCCGCGCTGCA 1106
DB 86 ArgAlaGlnArgAlaGlnArgAla 93

RESULT 11
US-09-709-188-10
; Sequence 10, Application US/09709188
; Patent No. 644137
; GENERAL INFORMATION:
; APPLICANT: Davis et al.
; TITLE OF INVENTION: Expressed Ligand - Vascular Intercellular Signaling Molecule
; FILE REFERENCE: REG 333-Z
; CURRENT APPLICATION NUMBER: US/09/709,188
; EARLIER FILING DATE: 2000-11-09
; EARLIER APPLICATION NUMBER: 08/740,223
; PRIOR FILING DATE: 1996-10-25
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-709-188-10

Alignment Scores:
Pred. No.: 44.4
Score: 8.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 1.20%
DB: 4
Length: 509
Matches: 8
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x US-09-709-188-10 (1-509)
QY 1083 CGTCCCGAGCGTCCCGCGCTGCA 1106
DB 86 ArgAlaGlnArgAlaGlnArgAla 93

RESULT 12
US-09-252-991a-18691
; Sequence 18691, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenstein et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
```

```

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18691
; LENGTH: 511
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-18691

Alignment Scores:
Pred. No.: 44.4
Score: 8.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 1.20%
DB: 4
Matches: 511
Conservative: 8
Mismatch: 0
Indels: 0
Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-09-252-991A-18691 (1-511)
QY 434 TCCCTACTGCGATTTCTAGCCTT 457
Db 406 SerLeuProAlaPheLeuSerLeu 413

RESULT 13
US-08-477-460B-4
; Sequence 4, Application US/08477460B
; Patent No. 6034223
; GENERAL INFORMATION:
; APPLICANT: Progenics Pharmaceuticals, Inc.
; TITLE OF INVENTION: NON-PEPTIDYL MOIETY-CONJUGATED
; NUMBER OF SEQUENCES: 9
; SOFTWARE: Patent In Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,460B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/927,931
; FILING DATE: 07-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 41215-A-PCT/JPM/AJM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 977-9809
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 530 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
; ORIGINAL SOURCE:
; COMPUTER READABLE FORM:

; ORGANISM: homo sapien
; CELL TYPE: Lymphocyte
; US-08-477-460B-4

Alignment Scores:
Pred. No.: 44.2
Score: 8.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 1.20%
DB: 3
Matches: 530
Conservative: 8
Mismatch: 0
Indels: 0
Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-08-477-460B-4 (1-530)
QY 1819 GACAGAGCTGACTCGGCGGAGC 1842
Db 78 AspArgAlaAspSerArgArgSer 85

RESULT 14
US-08-379-516-4
; Sequence 4, Application US/08379516
; Patent No. 6083478
; GENERAL INFORMATION:
; APPLICANT: Allaway, Graham P.
; APPLICANT: Madden, Paul J.
; TITLE OF INVENTION: No. 6083478-Peptidyl Moiety-Conjugated CD4-Gamma2 and CD4-IgG2
; FILE REFERENCE: 41215-A-PCT-US
; CURRENT APPLICATION NUMBER: US/08/379,516
; CURRENT FILING DATE: 1996-06-10
; EARLIER APPLICATION NUMBER: PCT/US93/07422
; EARLIER FILING DATE: 1993-08-06
; EARLIER APPLICATION NUMBER: 07/927,931
; EARLIER FILING DATE: 1992-08-07
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 4
; LENGTH: 530
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-08-379-516-4

Alignment Scores:
Pred. No.: 44.2
Score: 8.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 1.20%
DB: 3
Matches: 530
Conservative: 8
Mismatch: 0
Indels: 0
Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-08-379-516-4 (1-530)
QY 1819 GACAGAGCTGACTCGGCGGAGC 1842
Db 78 AspArgAlaAspSerArgArgSer 85

RESULT 15
US-09-329-916-4
; Sequence 4, Application US/09329916
; Patent No. 6177549
; GENERAL INFORMATION:
; APPLICANT: Progenics Pharmaceuticals, Inc.
; TITLE OF INVENTION: NON-PEPTIDYL MOIETY-CONJUGATED
; NUMBER OF SEQUENCES: 9
; SOFTWARE: Patent In Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US 07/927,931
; FILING DATE: 07-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 41215-A-PCT/JPM/AJM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 977-9809
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 530 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
; ORIGINAL SOURCE:
; COMPUTER READABLE FORM:
```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Releasee #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/329,916
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/477,460
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: US 07/927,931
FILING DATE: 07-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 41215-A-PCT/JPW/AJM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 977-9809
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 530 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: homo sapien
CELL TYPE: lymphocyte
US-09-329-916-4

Alignment Scores:
Pred. No.: 44.2 Length: 530
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 3 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-09-329-916-4 (1-530)

QY 1819 GACAGAGCTGACTCGCGCGGAGC 1842

DB 78 AsparGAlaAspSerArgArgSer 85

RESULT 16
US-08-485-372A-4
Sequence 4, Application US/08485372A
Patent No. 6187748
GENERAL INFORMATION:
APPLICANT: Beaudry, Gary A.
APPLICANT: Maddon, Paul J.
TITLE OF INVENTION: CD4-GAMMA2 CD4-IgG2 CHIMERAS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Releasee #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,372A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/476,227
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 37690-II-A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 530 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: homo sapien
CELL TYPE: lymphocyte
US-08-485-372A-4

Alignment Scores:
Pred. No.: 44.2 Length: 530
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 3 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-08-485-372A-4 (1-530)

QY 1819 GACAGAGCTGACTCGCGCGGAGC 1842

DB 78 AsparGAlaAspSerArgArgSer 85

RESULT 17
US-09-409-006A-4
Sequence 4, Application US/09409006A
Patent No. 6342586
GENERAL INFORMATION:
APPLICANT: Progenice Pharmaceuticals, Inc.
TITLE OF INVENTION: NON-PEPTIDYL MOIETY-CONJUGATED
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Releasee #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/409,006A
FILING DATE: 29-SEP-1999
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/927,931
FILING DATE: 07-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 41215-A-PCT/JPW/AJM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 977-9809
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
LENGTH: 530 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: unknown
ORIGINAL SOURCE: cDNA
ORGANISM: homo sapien
CELL TYPE: lymphocyte
US-09-409-006A-4

Alignment Scores:

Pred. No.:	44.2	Length:	530
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.20%	Indels:	0
DB:	4	Gaps:	0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-09-409-006A-4 (1-530)

QY 1819 GACAGAGCTGACTCGCGCGAGC 1842

Db 78 AsparGAlaaspSerArgArgSer 85

RESULT 18

US-08-484-681-4
Sequence 4, Application US/08484681
Patent No. 6451313
GENERAL INFORMATION:
APPLICANT: Beaudry, Gary A.
TITLE OF INVENTION: CD4-GAMMA2 CD4-IgG2 CHIMERAS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,681
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 37690-11-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 530 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: homo sapien
CELL TYPE: lymphocyte
US-08-484-681-4

Alignment Scores:

Pred. No.:	44.2	Length:	530
Score:	8.00	Matches:	8

Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 1.20%
DB: 4
Conservative: 0
Mismatches: 0
Indels: 0
Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-08-484-681-4 (1-530)

QY 1819 GACAGAGCTGACTCGCGCGAGC 1842

Db 78 AsparGAlaaspSerArgArgSer 85

RESULT 19

PCT-US93-07422-4
Sequence 4, Application PC/TUS9307422
GENERAL INFORMATION:
APPLICANT: Progenics Pharmaceuticals, Inc.
TITLE OF INVENTION: NON-PEPTIDYL MOLETY-CONJUGATED
TITLE OF INVENTION: CD4-GAMMA2 AND CD4-IgG2 IMMUNOCONJUGATES, AND USES THEREOF
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/07422
FILING DATE: 19930806
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/927,931
FILING DATE: 07-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 41215-A-PCT/JPW/JJM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 977-9809
TELEX: 422523 COOP UT
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 530 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: unknown
ORIGINAL SOURCE:
ORGANISM: homo sapien
CELL TYPE: lymphocyte
PCT-US93-07422-4

Alignment Scores:

Pred. No.:	44.2	Length:	530
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.20%	Indels:	0
DB:	5	Gaps:	0

US-10-029-345A-108_COPY_532_2532 (1-2001) x PCT-US93-07422-4 (1-530)

QY 1819 GACAGAGCTGACTCGCGCGAGC 1842

Db 78 AsparGAlaaspSerArgArgSer 85

RESULT 20

US-08-417-495-6
; Sequence 6, Application US/08417495
; Patent No. 5843728
; GENERAL INFORMATION:
; APPLICANT: Seed, Brian et al.
; TITLE OF INVENTION: Redirection of Cellular Immunity by
; TITLE OF INVENTION: Receptor
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: Wordperfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/417,495
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/203,866
; FILING DATE:
; APPLICATION NUMBER: US/07/847,566
; FILING DATE:
; APPLICATION NUMBER: 07/665,961
; FILING DATE: March 7, 1991
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/119002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 532 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-417-495-6

Alignment Scores:
Pred. No.: 44.2 Length: 532
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 2 Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x US-08-417-495-6 (1-532)
QY 1819 GACAGAGTACTGCGCGGAGC 1842
Db 78 AspaRgAlaSpSerRgArGser 85

RESULT 21
US-08-284-391B-6
; Sequence 6, Application US/08284391B
; Patent No. 5851828
; GENERAL INFORMATION:
; APPLICANT: Seed, Brian
; APPLICANT: Banapour, Babak
; APPLICANT: Romeo, Charles
; APPLICANT: Kolanus, Waldemar
; TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
; TITLE OF INVENTION: CELLS BY CHIMERIC CD4 RECEPTOR- BEARING CELLS

NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Elbing LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/284,391B
; FILING DATE: 02-AUG-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/195,395
; FILING DATE: 14-FEB-1994
; APPLICATION NUMBER: 07/847,566
; FILING DATE: 06-MAR-1992
; APPLICATION NUMBER: 07/665,961
; FILING DATE: 07-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Elbing, Karen L.
; REGISTRATION NUMBER: 35,238
; REFERENCE/DOCKET NUMBER: 00786/247001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-428-0200
; TELEFAX: 617-428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 532 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-284-391B-6

Alignment Scores:
Pred. No.: 44.2 Length: 532
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 2 Gaps: 0

US-10-029-345a-108_COPY_532_2532 (1-2001) x US-08-284-391B-6 (1-532)
QY 1819 GACAGAGTACTGCGCGGAGC 1842
Db 78 AspaRgAlaSpSerRgArGser 85

RESULT 22
US-09-218-950-6
; Sequence 6, Application US/09218950
; Patent No. 6284240
; GENERAL INFORMATION:
; APPLICANT: Seed, Brian
; APPLICANT: Banapour, Babak
; APPLICANT: Romeo, Charles
; APPLICANT: Kolanus, Waldemar
; TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
; TITLE OF INVENTION: CELLS BY CHIMERIC CD4 RECEPTOR- BEARING CELLS
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Elbing LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA

21P: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/218,950
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/284,391
FILING DATE: 02-AUG-1994
APPLICATION NUMBER: 08/195,395
FILING DATE: 14-FEB-1994
APPLICATION NUMBER: 07/847,566
FILING DATE: 06-MAR-1992
APPLICATION NUMBER: 07/665,961
FILING DATE: 07-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Elbing, Karen L.
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/247001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-218-950-6
Alignment Scores:
Pred. No.: 44.2
Score: 8.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 1.20%
Matches: 532
Conservative: 8
Mismatch: 0
Indels: 0
Gaps: 0
US-10-029-345A-108_COPY_532_2532 (1-2001) x US-09-218-950-6 (1-532)
QY 1819 GACAGAGCTGACTCGCGCGGAGC 1842
DB 78 AsparGAlaAspSerArgArgSer 85
RESULT 23
PCT-US92-01785-6
Sequence 6, Application PC/TUS9201785
GENERAL INFORMATION:
APPLICANT: The General Hospital Corporation
TITLE OF INVENTION: Redirection of Cellular Immunity by Receptor
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: Wordperfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/01785
FILING DATE: 19920306

CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/665,961
FILING DATE: March 7, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/119002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US92-01785-6
Alignment Scores:
Pred. No.: 44.2
Score: 8.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 1.20%
Matches: 532
Conservative: 8
Mismatch: 0
Indels: 0
Gaps: 0
US-10-029-345A-108_COPY_532_2532 (1-2001) x PCT-US92-01785-6 (1-532)
QY 1819 GACAGAGCTGACTCGCGCGGAGC 1842
DB 78 AsparGAlaAspSerArgArgSer 85
RESULT 24
PCT-US95-00454-6
Sequence 6, Application PC/TUS9500454
GENERAL INFORMATION:
APPLICANT: Seed, Brian et al.
TITLE OF INVENTION: Targeted Cytolysis of HIV-Infected
TITLE OF INVENTION: Cells by Chimeric CD4 Receptor-
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: Wordperfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/00454
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/847,566
FILING DATE: March 6, 1992
CLASSIFICATION:
APPLICATION NUMBER: 07/665,961
FILING DATE: March 7, 1991
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/247001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070

TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-00454-6

Alignment Scores:
Pred. No.: 44.2 Length: 532
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x PCT-US95-00454-6 (1-532)

QY 1819 GACAGAGCTGACTCGCGCGGAGC 1842
Db 78 AspArgAlaAspSerArgSer 85

RESULT 25
US-08-417-495-4
Sequence 4, Application US/08417495
Patent No. 5843728
GENERAL INFORMATION:
APPLICANT: Seed, Brian et al.
TITLE OF INVENTION: Redirection of Cellular Immunity by
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 555X
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: Wordperfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/417,495
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/203,866
FILING DATE:
APPLICATION NUMBER: US/07/847,566
FILING DATE:
APPLICATION NUMBER: 07/665,961
FILING DATE: March 7, 1991
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/119002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 575 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-417-495-4

Alignment Scores:
Pred. No.: 43.9 Length: 575
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-08-417-495-4 (1-575)

QY 1819 GACAGAGCTGACTCGCGCGGAGC 1842
Db 78 AspArgAlaAspSerArgSer 85

RESULT 26
US-08-284-391B-4
Sequence 4, Application US/08284391B
Patent No. 5851828
GENERAL INFORMATION:
APPLICANT: Seed, Brian
APPLICANT: Banapour, Babak
APPLICANT: Romeo, Charles
APPLICANT: Kolanus, Waldemar
TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
TITLE OF INVENTION: CELLS BY CHIMERIC CD4 RECEPTOR-BEARING CELLS
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Clark & Elbing LLP
STREET: 176 Federal Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: PastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/284,391B
FILING DATE: 02-AUG-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/195,395
FILING DATE: 14-FEB-1994
APPLICATION NUMBER: 07/847,566
FILING DATE: 06-MAR-1992
APPLICATION NUMBER: 07/665,961
FILING DATE: 07-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Elbing, Karen L.
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/247001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 575 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-284-391B-4

Alignment Scores:
Pred. No.: 43.9 Length: 575
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-08-284-391B-4 (1-575)

OY 1819 GACAGAGCTGACTCGCGGCGAGC 1842

DB 78 AspaRgAlaSpSerArGArSer 85

RESULT 27

US-09-218-950-4

Sequence 4, Application US/09218950

Patent No. 6284240

GENERAL INFORMATION:

APPLICANT: Seed, Brian

APPLICANT: Banapour, Babak

APPLICANT: Romeo, Charles

APPLICANT: Kolanus, Waldemar

TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED

TITLE OF INVENTION: CELLS BY CHIMERIC CD4 RECEPTOR-BEARING CELLS

NUMBER OF SEQUENCES: 53

CORRESPONDENCE ADDRESS:

ADDRESSEE: Clark & Elbing LLP

STREET: 176 Federal Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02110

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/218,950

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/284,391

FILING DATE: 02-AUG-1994

APPLICATION NUMBER: 08/195,395

FILING DATE: 14-FEB-1994

APPLICATION NUMBER: 07/847,566

FILING DATE: 06-MAR-1992

APPLICATION NUMBER: 07/665,961

FILING DATE: 07-MAR-1991

ATTORNEY/AGENT INFORMATION:

NAME: Elbing, Karen L.

REGISTRATION NUMBER: 35,238

REFERENCE/DOCKET NUMBER: 00786/247001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-428-0200

TELEFAX: 617-428-7045

TELEX:

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 575 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-218-950-4

Alignment Scores:

Pred. No.: 43.9 Length: 575

Score: 8.00 Matches: 8

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 1.20% Indels: 0

DB: 3 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-09-218-950-4 (1-575)

OY 1819 GACAGAGCTGACTCGCGGCGAGC 1842

DB: 78 AspaRgAlaSpSerArGArSer 85

DB 78 AspaRgAlaSpSerArGArSer 85

RESULT 28

PCT-US92-01785-4

Sequence 4, Application PC/TUS9201785

GENERAL INFORMATION:

APPLICANT: The General Hospital Corporation

TITLE OF INVENTION: Redirection of Cellular Immunity by Receptor

TITLE OF INVENTION: Chimeras

NUMBER OF SEQUENCES: 27

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson

STREET: 225 Franklin Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

COMPUTER: IBM PS/2 Model 502 or 55SX

OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)

SOFTWARE: Wordperfect (Version 5.0)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US92/01785

FILING DATE: 19920306

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/665,961

FILING DATE: March 7, 1991

ATTORNEY/AGENT INFORMATION:

NAME: Clark, Paul T.

REGISTRATION NUMBER: 30,162

REFERENCE/DOCKET NUMBER: 00786/119002

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 542-5070

TELEFAX: (617) 542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 575 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

PCT-US92-01785-4

Alignment Scores:

Pred. No.: 43.9 Length: 575

Score: 8.00 Matches: 8

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 1.20% Indels: 0

DB: 5 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x PCT-US92-01785-4 (1-575)

OY 1819 GACAGAGCTGACTCGCGGCGAGC 1842

DB 78 AspaRgAlaSpSerArGArSer 85

RESULT 29

PCT-US95-00454-4

Sequence 4, Application PC/TUS9500454

GENERAL INFORMATION:

APPLICANT: Seed, Brian et al.

TITLE OF INVENTION: Targeted Cytolysis of HIV-Infected

TITLE OF INVENTION: Cells by Chimeric CD4 Receptor-

NUMBER OF SEQUENCES: 27

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson

STREET: 225 Franklin Street

CITY: Boston

STATE: MA

COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: Mordperfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/00454
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/847,566
FILING DATE: March 6, 1992
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/665,961
FILING DATE: March 7, 1991
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/247001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 575 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-00454-4

Alignment Scores:
Pred. No.: 43.9 Length: 575
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x PCT-US95-00454-4 (1-575)

QY 1819 GACAGCTGACTCGCGCGGAGC 1842
|||||
Db 78 AspArgLalaPseSerArgSer 85

RESULT 30
US-09-252-991A-18124
Sequence 18124, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 18124
LENGTH: 592
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18124

Alignment Scores:
Pred. No.: 43.7 Length: 592

Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0
DB: Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-09-252-991A-18124 (1-592)

QY 1113 CGACGGCTGACGCTGCGACGCT 1090
|||||
Db 302 ArgArgLeuValIleAlaIleAla 309

RESULT 31
US-08-746-283-32
Sequence 32, Application US/08746283
Patent No. 5834280
GENERAL INFORMATION:
APPLICANT: Okenbolil, Karen M.
APPLICANT: St. Joan O.
APPLICANT: Aagaard, Jesper
TITLE OF INVENTION: Glucose Oxidases
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSER: No. 58342800 No. 5834280disk of No. 5834280th America
STREET: 405 Lexington Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/746,283
FILING DATE: 07-NOV-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Agtib, Dr. Cheryl H.
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 4158.214-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 867-0123
TELEFAX: (212) 878-9655
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 604 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-746-283-32

Alignment Scores:
Pred. No.: 43.6 Length: 604
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-08-746-283-32 (1-604)

QY 386 AGTTCCTGCTGTTGTTCCCTGCGC 409
|||||
Db 7 SerSerLeuValValSerLeuAla 14

RESULT 32
US-08-746-257A-30
Sequence 30, Application US/08746257A
Patent No. 5879921
GENERAL INFORMATION:
APPLICANT: Cherry, Joel

```

;
; APPLICANT: Berka, Randy
; APPLICANT: Halkier, Torben
; TITLE OF INVENTION: Glucose Oxidases
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 58799210 No. 5879921disk of No. 5879921th America
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/746,257A
; FILING DATE: 07-NOV-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Starnes, Robert L.
; REGISTRATION NUMBER: 41,324
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
;
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 604 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-08-746-257A-30
;
; Alignment Scores:
; Pred. No.: 43.6 Length: 604
; Score: 8.00 Matches: 8
; Percent Similarity: 100.00% Conservative: 0
; Best Local Similarity: 100.00% Mismatches: 0
; Query Match: 1.20% Indels: 0
; DB: 2 Gaps: 0
;
; US-10-029-345A-108_COPY_532_2532 (1-2001) x US-08-746-257A-30 (1-604)
;
; Oy 386 AGTTCCTCGTTGTTCCCTGACC 409
;
; Db 7 SerSerLeuValValSerLeuAla 14
;
; RESULT 33
; US-08-333-802-2
; Sequence 2, Application US/08333802
; Patent No. 5516671
;
; GENERAL INFORMATION:
; APPLICANT: Lawrence, Ellen B.
; APPLICANT: Levine, Elaine B.
; APPLICANT: Shah, Dilipkumar M.
; TITLE OF INVENTION: Method of Controlling Plant Pathogens
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Grace L. Bonner, Monsanto Company, BB4F
; STREET: 700 Chesterfield Parkway No. 5516671th
; CITY: St. Louis
; STATE: Missouri
; COUNTRY: USA
; ZIP: 63198
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

```

;
; APPLICATION NUMBER: US/08/333,802
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161041
; FILING DATE: 24-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Bonner, Grace L.
; REGISTRATION NUMBER: 32,963
; REFERENCE/DOCKET NUMBER: 38-21(10663)A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (314)537-7286
; TELEFAX: (314)537-6047
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 605 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
;
; US-08-333-802-2
;
; Alignment Scores:
; Pred. No.: 43.6 Length: 605
; Score: 8.00 Matches: 8
; Percent Similarity: 100.00% Conservative: 0
; Best Local Similarity: 100.00% Mismatches: 0
; Query Match: 1.20% Indels: 0
; DB: 1 Gaps: 0
;
; US-10-029-345A-108_COPY_532_2532 (1-2001) x US-08-333-802-2 (1-605)
;
; Oy 386 AGTTCCTCGTTGTTCCCTGACC 409
;
; Db 7 SerSerLeuValValSerLeuAla 14
;
; RESULT 34
; US-09-347-878-46
; Sequence 46, Application US/09347878C
; Patent No. 6376210
; GENERAL INFORMATION:
; APPLICANT: Yuan, Chong
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ASSAYING ANALYTES
; FILE REFERENCE: 25885-1651
; CURRENT APPLICATION NUMBER: US/09/347,878C
; CURRENT FILING DATE: 1999-07-06
; NUMBER OF SEQ ID NOS: 75
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 46
; TYPE: PRT
; LENGTH: 605
; ORGANISM: Aspergillus niger
;
; US-09-347-878-46
;
; Alignment Scores:
; Pred. No.: 43.6 Length: 605
; Score: 8.00 Matches: 8
; Percent Similarity: 100.00% Conservative: 0
; Best Local Similarity: 100.00% Mismatches: 0
; Query Match: 1.20% Indels: 0
; DB: 4 Gaps: 0
;
; US-10-029-345A-108_COPY_532_2532 (1-2001) x US-09-347-878-46 (1-605)
;
; Oy 386 AGTTCCTCGTTGTTCCCTGACC 409
;
; Db 7 SerSerLeuValValSerLeuAla 14
;
; RESULT 35
; US-09-252-991A-26137
; Sequence 26137, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.

```

```
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252.991A
/ CURRENT FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074.788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094.190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 26137
/ LENGTH: 607
/ TYPE: PRT
/ ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-26137

Alignment Scores:
Pred. No.: 43.6 Length: 607
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0
Gaps: 0
DB: 4

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-09-252-991A-26137 (1-607)
QY 1622 CAGCCCTTAAGCCCGAGCCAGCA 1599
DB 193 GlnProLeuArgProArgProAla 200

RESULT 36
US-09-252-991A-28230
/ Sequence 28230; Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252.991A
/ CURRENT FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074.788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094.190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 28230
/ LENGTH: 611
/ TYPE: PRT
/ ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28230

Alignment Scores:
Pred. No.: 43.6 Length: 611
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
Gaps: 0
DB: 4

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-09-252-991A-28230 (1-611)
QY 1158 CCTGTCGAGACAGGCTGGAAGA 1181
DB 360 ProValArgArgGlnAlaGlyArg 367

RESULT 37
US-08-746-283-31
/ Sequence 31; Application US/08746283
/ Patent No. 5834280
/ GENERAL INFORMATION:
/ APPLICANT: Oxenboll, Karen M.
/ APPLICANT: Si, Joan Q.
```

```
/ APPLICANT: Aagaard, Jesper
/ TITLE OF INVENTION: Glucose Oxidases
/ NUMBER OF SEQUENCES: 32
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: No. 58342800 No. 5834280disk of No. 5834280th America
/ STREET: 405 Lexington Avenue
/ CITY: New York
/ STATE: New York
/ COUNTRY: USA
/ ZIP: 10174
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/ SOFTWARE: FASTSEQ for Windows Version 2.0
/ CURRENT APPLICATION DATA:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Agtis, Dr. Cheryl H.
/ REGISTRATION NUMBER: 34,086
/ REFERENCE/DOCKET NUMBER: 4158.214-US
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (212) 867-0123
/ TELEFAX: (212) 878-9655
/ INFORMATION FOR SEQ ID NO: 31:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 612 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: No. 5834280e
US-08-746-283-31

Alignment Scores:
Pred. No.: 43.6 Length: 612
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0
Gaps: 0
DB: 2

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-08-746-283-31 (1-612)
QY 1593 CGTAGGTCGCTGCTGCTGGA 1570
DB 307 ArgGluValLeuLeuAlaGlyGly 314

RESULT 38
US-08-746-257A-29
/ Sequence 29; Application US/08746257A
/ Patent No. 5879921
/ GENERAL INFORMATION:
/ APPLICANT: Cherry, Joel
/ APPLICANT: Berka, Randy
/ APPLICANT: Halkier, Torben
/ TITLE OF INVENTION: Glucose Oxidases
/ NUMBER OF SEQUENCES: 30
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: No. 58799210 No. 5879921disk of No. 5879921ch America
/ STREET: 405 Lexington Avenue
/ CITY: New York
/ STATE: NY
/ COUNTRY: USA
/ ZIP: 10174
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: DOS
/ SOFTWARE: FASTSEQ for Windows Version 2.0
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/746.257A
```

FILING DATE: 07-NOV-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Starnes, Robert L.
REGISTRATION NUMBER: 41,334
REFERENCE/DOCKET NUMBER: 4700.000-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 612 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-746-257A-29
Alignment Scores:
Pred. No.: 43.6 Length: 612
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0
DB: 2 Gaps: 0
US-10-029-345A-108_COPY_532_2532 (1-2001) x US-08-746-257A-29 (1-612)
QY 1593 CGTGAGTGTCTGCTGGCTGTGGA 1570
Db 307 ArgGluValIleuLeuAlaGlyGly 314
RESULT 39
US-08-746-283-1
Sequence 1, Application US/08746283
Patent No. 5834280
GENERAL INFORMATION:
APPLICANT: Okenboll, Karen M.
APPLICANT: Si, Joan O.
APPLICANT: Aagaard, Jesper
TITLE OF INVENTION: Glucose Oxidases
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 58342800 No. 5834280disk of No. 5834280th America
STREET: 405 Lexington Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/746,283
FILING DATE: 07-NOV-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Agtis, Dr. Cheryl H.
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 4158.214-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 867-0123
TELEFAX: (212) 878-9655
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 613 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-746-283-1

Alignment Scores:
Pred. No.: 43.6 Length: 613
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0
DB: 2 Gaps: 0
US-10-029-345A-108_COPY_532_2532 (1-2001) x US-08-746-283-1 (1-613)
QY 1593 CGTGAGTGTCTGCTGGCTGTGGA 1570
Db 308 ArgGluValIleuLeuAlaGlyGly 315
RESULT 40
US-08-746-283-3
Sequence 3, Application US/08746283
Patent No. 5834280
GENERAL INFORMATION:
APPLICANT: Okenboll, Karen M.
APPLICANT: Si, Joan O.
APPLICANT: Aagaard, Jesper
TITLE OF INVENTION: Glucose Oxidases
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 58342800 No. 5834280disk of No. 5834280th America
STREET: 405 Lexington Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/746,283
FILING DATE: 07-NOV-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Agtis, Dr. Cheryl H.
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 4158.214-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 867-0123
TELEFAX: (212) 878-9655
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 613 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-746-283-3
Alignment Scores:
Pred. No.: 43.6 Length: 613
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0
DB: 2 Gaps: 0
US-10-029-345A-108_COPY_532_2532 (1-2001) x US-08-746-283-3 (1-613)
QY 1593 CGTGAGTGTCTGCTGGCTGTGGA 1570
Db 308 ArgGluValIleuLeuAlaGlyGly 315
RESULT 41
US-08-746-283-5

Sequence 5, Application US/08746283
Patent No. 5834280
GENERAL INFORMATION:
APPLICANT: Oxendoll, Karen M.
APPLICANT: Si, Joan Q.
APPLICANT: Aagaard, Jesper
TITLE OF INVENTION: Glucose Oxidases
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 58342800 No. 5834280disk of No. 5834280th America
STREET: 405 Lexington Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/746,283
FILING DATE: 07-NOV-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Agria, Dr. Cheryl H.
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 4158,214-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 867-0123
TELEFAX: (212) 878-9655
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 613 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-746-283-5

Alignment Scores:
Pred. No.: 43.6 Length: 613
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0
Gaps: 0
DB: 2

US-10-029-345a-108_COPY_532_2532 (1-2001) x US-08-746-283-5 (1-613)

QY 1593 CGTGAGTGCTGCTGCTGTGGA 1570
Db 308 ArgGluValLeuLeuAlaGlyGly 315

RESULT 42
US-08-746-257A-1
Sequence 1, Application US/08746257A
Patent No. 5879921
GENERAL INFORMATION:
APPLICANT: Cherry, Joel
APPLICANT: Berka, Randy
APPLICANT: Halkier, Torben
TITLE OF INVENTION: Glucose Oxidases
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 58799210 No. 5879921disk of No. 5879921th America
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/746,257A
FILING DATE: 07-NOV-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Starnes, Robert L.
REGISTRATION NUMBER: 41,324
REFERENCE/DOCKET NUMBER: 4700,000-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 613 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-746-257A-1

Alignment Scores:
Pred. No.: 43.6 Length: 613
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0
Gaps: 0
DB: 2

US-10-029-345a-108_COPY_532_2532 (1-2001) x US-08-746-257A-1 (1-613)

QY 1593 CGTGAGTGCTGCTGCTGTGGA 1570
Db 308 ArgGluValLeuLeuAlaGlyGly 315

RESULT 43
US-09-347-878-48
Sequence 48, Application US/09347878C
Patent No. 6376210
GENERAL INFORMATION:
APPLICANT: Yuan, Cheng
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ASSAYING ANALYTES
FILE REFERENCE: 25885-1651
CURRENT APPLICATION NUMBER: US/09/347,878C
CURRENT FILING DATE: 1999-07-06
NUMBER OF SEQ ID NOS: 75
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 48
LENGTH: 613
TYPE: PRT
ORGANISM: Cladosporium oxysporum
FEATURE:
OTHER INFORMATION: Glucose oxidase protein sequence
PUBLICATION INFORMATION:
PATENT DOCUMENT NUMBER: 08/746,257
PATENT FILING DATE: 1996-11-07
PUBLICATION DATE: 1999-03-09
US-09-347-878-48

Alignment Scores:
Pred. No.: 43.6 Length: 613
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.24% Indels: 0
Gaps: 0
DB: 4

US-10-029-345a-108_COPY_532_2532 (1-2001) x US-09-347-878-48 (1-613)

QY 1593 CGTGAGTGCTGCTGCTGTGGA 1570

```
Db          308 ArgGlValLeuLeuLagLygLy 315
|||||
RESULT 44
US-08-472-888A-6
; Sequence 6, Application US/08472888A
; Patent No. 6613746
; GENERAL INFORMATION:
; APPLICANT: Seed, Brian
; APPLICANT: Walz, Gerd
; TITLE OF INVENTION: AGP-ANTIBODY FUSION PROTEINS
; TITLE OF INVENTION: AND RELATED MOLECULES AND METHODS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Clark & Elbing LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,888A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/618,314
; FILING DATE: 23-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Elbing, Karen L.
; REGISTRATION NUMBER: 35,238
; REFERENCE/DOCKET NUMBER: 00786/258001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-428-0200
; TELEFAX: 617-428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 630 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-472-888A-6
Alignment Scores:
Pred. No.: 43.4 Length: 630
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
Gaps: 0
DB: 4
US-10-029-345A-108_COPY_532_2532 (1-2001) x US-08-472-888A-6 (1-630)
QY          1819 GACGAGCTGACTCGGCGCGGAC 1842
|||||
Db          78 AspArgAlaAspSerArgArgSer 85
|||||
RESULT 45
US-09-591-095-6
; Sequence 6, Application US/09591095
; Patent No. 6489461
; GENERAL INFORMATION:
; APPLICANT: Froman, B. & Dehesh, K.
; TITLE OF INVENTION: Nucleic Acid Sequences encoding Proteins
; TITLE OF INVENTION: Involved in Fatty Acid Beta-Oxidation and Methods of Use.
; FILE REFERENCE: 17137/01/US
; CURRENT APPLICATION NUMBER: US/09/591,095
```

```
; CURRENT FILING DATE: 2000-06-08
; EARLIER APPLICATION NUMBER: US 60/138,162
; EARLIER FILING DATE: 1999-06-08
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 683
; TYPE: PRT
; ORGANISM: Arabidopsis sp.
US-09-591-095-6
Alignment Scores:
Pred. No.: 43.1 Length: 683
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
Gaps: 0
DB: 4
US-10-029-345A-108_COPY_532_2532 (1-2001) x US-09-591-095-6 (1-683)
QY          596 AGCTGACTTATCCCGAGTCTC 619
|||||
Db          25 SerLeuThrLeuSerProSerIleu 32
|||||
RESULT 46
US-08-021-601-12
; Sequence 12, Application US/08021601
; Patent No. 5591631
; GENERAL INFORMATION:
; APPLICANT: Leppla, Stephen H.
; APPLICANT: Klimpel, Kurt R.
; APPLICANT: Nichols, Peter J.
; APPLICANT: Atora, Naveen
; APPLICANT: Singh, Yogendra
; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
; TITLE OF INVENTION: RELATED METHODS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Needle & Rosenberg, P.C.
; STREET: 133 Carnegie Way, Suite 400
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/021,601
; FILING DATE: 19930212
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.057
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404/688-0770
; TELEFAX: 404/688-9880
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 903 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-021-601-12
Alignment Scores:
Pred. No.: 41.9 Length: 903
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
```


Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 1 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-08-021-601-12 (1-903)

QY 1819 GACAGAGCTGACTCGCGCGGAGC 1842

DB 778 AEPARGAIAASPERRARGARGSER 785

RESULT 47

US-08-082-849B-12

Sequence 12, Application US/08082849B

Patent No. 5677274

GENERAL INFORMATION:

APPLICANT: Leppia, Stephen H.

APPLICANT: Klimpel, Kurt R.

APPLICANT: Airola, Naveen

APPLICANT: Singh, Yogendra

APPLICANT: Nichols, Peter J.

TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and

TITLE OF INVENTION: Related Methods

NUMBER OF SEQUENCES: 35

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, Eighth Floor

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/082,849B

FILING DATE: 25-JUN-1993

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/021,601

FILING DATE: 12-FEB-1993

ATTORNEY/AGENT INFORMATION:

NAME: Weber, Kenneth A.

REGISTRATION NUMBER: 31,677

REFERENCE/DOCKET NUMBER: 15280-161-1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 903 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-082-849B-12

Alignment Scores:

Pred. No.: 41.9 Length: 903

Score: 8.00 Matches: 8

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 1.20% Indels: 0

DB: 1 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-08-082-849B-12 (1-903)

QY 1819 GACAGAGCTGACTCGCGCGGAGC 1842

DB 778 AEPARGAIAASPERRARGARGSER 785

RESULT 48

PCT-US94-01624-12

Sequence 12, Application PC/TUS9401624

GENERAL INFORMATION:

APPLICANT: Leppia, Stephen H.

APPLICANT: Klimpel, Kurt R.

APPLICANT: Airola, Naveen

APPLICANT: Singh, Yogendra

APPLICANT: Nichols, Peter J.

TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND

TITLE OF INVENTION: RELATED METHODS

NUMBER OF SEQUENCES: 31

CORRESPONDENCE ADDRESS:

ADDRESSEE: TOWNSEND and TOWNSEND KHOURIE and CREW

STREET: Stewart Street Tower, 20th Floor, One Market

STREET: Plaza

CITY: San Francisco

STATE: CA

COUNTRY: USA

ZIP: 94105

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US94/01624

FILING DATE: June 25, 1993

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Weber, Kenneth A.

REGISTRATION NUMBER: 31,677

REFERENCE/DOCKET NUMBER: 15280-115

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 543-9600

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 903 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

PCT-US94-01624-12

Alignment Scores:

Pred. No.: 41.9 Length: 903

Score: 8.00 Matches: 8

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 1.20% Indels: 0

DB: 5 Gaps: 0

US-10-029-345A-108_COPY_532_2532 (1-2001) x PCT-US94-01624-12 (1-903)

QY 1819 GACAGAGCTGACTCGCGCGGAGC 1842

DB 778 AEPARGAIAASPERRARGARGSER 785

RESULT 49

US-09-252-991A-33037

Sequence 33037, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NOCLETIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

PRIOR FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 33037

; LENGTH: 935
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-33037

Alignment Scores:

Pred. No.:	41.7	Length:	935
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.20%	Indels:	0
DB:	4	Gaps:	0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-09-252-991A-33037 (1-935)

QY 1122 GGACAGCCCGCTGTGTACAGGCGCT 1145

DB 307 GlyGlnProAlaGlyThrGlyAla 314

RESULT 50

US-09-252-991A-27230
; Sequence 27230, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27230
; LENGTH: 953
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-27230

Alignment Scores:

Pred. No.:	41.6	Length:	953
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.24%	Indels:	0
DB:	4	Gaps:	0

US-10-029-345A-108_COPY_532_2532 (1-2001) x US-09-252-991A-27230 (1-953)

QY 1107 CTGCACGCTGGGCGACGCTGGGCAC 1084

DB 241 LeuHisAlaGlyHisAlaGlyHis 248

Search completed: June 21, 2004, 13:15:47
Job time : 62.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2004, 23:40:27 ; Search time 12406 Seconds

(without alignments)
6969.962 Million cell updates/sec

Title: US-10-029-345a-108_COPY_538_2532

Perfect score: 1995
Sequence: 1 atggcccatgagatgattgg.....tggaaatcatgaggtctcc 1995

Scoring table: OLIGO_NUC
Gapop 60.0, Gapext 60.0

Searched: 3470272 seqs, 2167151695 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1263658

Minimum DB seq length: 1419

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

```

1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_ov:*
5: gb_om:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_stb:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_mu:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_stb:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_hum:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rnd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_by:*
39: em_hgtg_hum:*
40: em_hgtg_mus:*
41: em_hgtg_other:*

```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1995	100.0	5450	6	AX482439	Sequence
2	1995	100.0	5450	6	AX482478	Sequence
3	1893	94.9	1998	6	AX260342	Sequence
4	1893	94.9	3059	6	AX278461	Sequence
5	1893	94.9	3496	6	AX441210	Sequence
6	1893	94.9	3521	9	AB052156	Sequence
7	1893	94.9	3544	6	AX260340	Sequence
8	1893	94.9	3566	9	AF506796	Sequence
9	1893	94.9	3766	6	AX374994	Sequence
10	1893	94.9	4790	6	BD171157	Novel gen
11	1893	94.9	4790	6	BD183422	Novel gen
12	1893	94.9	4790	6	AB051487	Sequence
13	1842	92.3	2732	6	AX180875	Sequence
14	1842	92.3	3104	6	AX405700	Sequence
15	1791	89.8	2102	6	AX713989	Sequence
16	1791	89.8	2102	6	AK055973	Sequence
17	1791	89.8	5111	6	AX482372	Sequence
18	1696	85.0	2807	6	BC031643	Sequence
19	1593	79.8	2118	6	AX099933	Sequence
20	1418	71.1	1935	9	AY038927	Homo sapi
21	1418	71.1	3332	6	AX441229	Sequence
22	1284	64.4	3284	9	BC042101	Homo sapi
23	1233	61.8	2071	6	AX921917	Sequence
24	1183	59.3	172206	9	AC007619	Sequence
25	1183	59.3	188344	2	AC131617	Homo sapi
26	1182	59.2	2200	6	AX921915	Sequence
27	788	39.5	1916	6	AX835305	Sequence
28	788	39.5	1916	9	AK098310	Homo sapi
29	143	7.2	5047	9	HSMB004384	Sequence
30	79	4.0	2756	6	AX482444	Sequence
31	79	4.0	4874	10	AB052157	Mus muscu
32	79	4.0	4943	10	BC059232	Mus muscu
33	79	4.0	4975	10	BC057321	Mus muscu
34	79	4.0	201474	2	AC126692	Mus muscu
35	79	4.0	236589	2	AC118035	Mus muscu
36	76	3.8	244605	2	AC097818	Rattus no
37	76	3.8	250782	2	AC128093	Rattus no
38	72	3.6	3861	10	AF345954	Mus muscu
39	72	3.6	4026	10	AF345953	Mus muscu
40	72	3.6	4827	10	AF345952	Mus muscu
41	72	3.6	4992	10	AF345951	Mus muscu
42	53	2.7	201474	2	AC126692	Mus muscu
43	40	2.0	247148	2	AC097776	Rattus no
44	38	1.9	242590	2	AC133722	Rattus no
45	22	1.1	84077	9	AL157371	Human DNA

ALIGNMENTS

```

RESULT 1
AX482439
LOCUS AX482439 5450 bp DNA linear PAT 16-AUG-2002
DEFINITION Sequence 108 from Patent WO02057460.
ACCESSION AX482439
VERSION AX482439.1 GI:22316984
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1
Toderud,C.G., Bol,D., Finger,J., Lee,J., Nelson,T., Schieven,G.,
Suchard,S., Banas,D., Bassolino,D., Feder,J., Kyslek,S.,
Mcatee,P., Mintier,G., Siemers,N., Jackson,D.G. and Ramamathan,C.

```

TITLE Polynucleotides encoding human phosphatases
JOURNAL Patent: WO 02057460-A 108 25-JUL-2002;
SOUTHB BRISTOL MYERS CO (US)

FEATURES
Source Location/Qualifiers

1..5450
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
538..2535
/note="unnamed protein product"

CDS

/codon_start=1
/protein_id="CAD44451.1"
/db_xref="GI:22316985"
/translation="MAHMTGTQIVTERLHSAKHYDIDDSQVYVYDDSSQVASSDSC
FLVYLKLEKSRNSVHLLAGRAEFERCPGLCEGKSTLVPTCISQPCLVANIGPT
RIIPNLVIGCORVLYNKEILIQNGIYVLYNASTYCPKDFI PESHFLRVNDSFCK
ILPWLDSVDFIERAKNSGCVLHVCLAGISRSATLAIAYIMKRMDSIDEAVRYE
KRPTISPNFNLGQLDYEKIKNOTGASGPKSKLHLKENEKPEVPAVSEGGKSE
TLPSPCADSATSEAGORPVHPASVPSVPSVOPSLIEDSLVLOALSGHLSDRLD
SNLKRSEFSDIKSVSASMAASLHGSSSPALHYKRSSTLDTGNLKCSPYOE
LSBOTPESTDKPEASIPKYLQTRPESDSKRLHSVRSSSGIAQRLSLPLHRSGS
VEDNYHTSPFLGISTSOQHLTKSAGLKLKGMHSDILAPOTSTSLTSSWTFATESHF
YSASATIGSASISVASCQLPTCGDQVSVRRORKPSDADRRSRWHERSPREKQFK
RRCQMEFGESIMSENRSBELKVGSSFSGSMELIEVS"

ORIGIN

Query Match 100.0%; Score 1995; DB 6; Length 5450;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1995; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGGCCCATGATGATGATGAACTCAATTTGTTACTGAGAGTGTGCTCTGTGGA	60
DB	538	ATGGCCCATGATGATGATGAACTCAATTTGTTACTGAGAGTGTGCTCTGTGGA	597
QY	61	AGTGAACGGAAAAAGCTGCTAATTGATAGCCGGCATTGTGAAATACATACCC	120
DB	598	AGTGAACGGAAAAAGCTGCTAATTGATAGCCGGCATTGTGAAATACATACCC	657
QY	121	CACATTTTGAAGCATTATATCACTGCTCCAGCTTATGAAAGGAAGTTGCAACG	180
DB	658	CACATTTTGAAGCATTATATCACTGCTCCAGCTTATGAAAGGAAGTTGCAACG	717
QY	181	GACAAAGTTAATACAGAGCTCATCCAGCATTCAGCAAAATTAAGTTGACATTT	240
DB	718	GACAAAGTTAATACAGAGCTCATCCAGCATTCAGCAAAATTAAGTTGACATTT	777
QY	241	TGCAATCAGAAAGTTAGTTACGATCAAAAGTCCCAAGATGTTCCCTCTCTTCA	300
DB	778	TGCAATCAGAAAGTTAGTTACGATCAAAAGTCCCAAGATGTTCCCTCTCTTCA	837
QY	301	GACTGTTTCTCACTGACTCTGAGTAACTGAGAAAGCTTCAACTCTGTACCGT	360
DB	838	GACTGTTTCTCACTGACTCTGAGTAACTGAGAAAGCTTCAACTCTGTACCGT	897
QY	361	CTTGCAAGTGGTGTCTGAGATTCTCTGCTGTTTCCCTGAGCTCTGTGAAGAAATCC	420
DB	898	CTTGCAAGTGGTGTCTGAGATTCTCTGCTGTTTCCCTGAGCTCTGTGAAGAAATCC	957
QY	421	ACTCTAGCCCTACCGCATTTCTCAGCCTTCTTACCTGTGGCCAACTTGGCCAAAC	480
DB	958	ACTCTAGCCCTACCGCATTTCTCAGCCTTCTTACCTGTGGCCAACTTGGCCAAAC	1017
QY	481	CGAATTTCTCCAACTTTATCTTGCTGCGCAGCAGAGATGCTCAACAAGAGCTGATA	540
DB	1018	CGAATTTCTCCAACTTTATCTTGCTGCGCAGCAGAGATGCTCAACAAGAGCTGATA	1077
QY	541	CAGCAGATGGGATTTGTTATGTTAAATGCACTATACCTGTCCAAAGCTGACTTT	600
DB	1078	CAGCAGATGGGATTTGTTATGTTAAATGCACTATACCTGTCCAAAGCTGACTTT	1137
QY	601	ATCCCGAGTCTCATTTCTGTGCTGTGCTGATGATGACGCTTTTGTGAAGAAATTTTG	660

DB	1138	ATCCCGAGTCTCATTTCTGTGCTGTGCTGATGATGACGCTTTGTGAAGAAATTTTG	1197
QY	661	CCGTGTTGGAACAATCAGTAGATTTTCATTGAGAAAGCAAAAGCCCTCAATGGATGTG	720
DB	1198	CCGTGTTGGAACAATCAGTAGATTTTCATTGAGAAAGCAAAAGCCCTCAATGGATGTG	1257
QY	721	CTTAGTCACTGTTTACGTGGGATCTCCGCTCGCCCAACCATGCTATCGCTACATG	780
DB	1258	CTTAGTCACTGTTTACGTGGGATCTCCGCTCGCCCAACCATGCTATCGCTACATG	1317
QY	781	AAGAGATGAGATGCTTTTATAGTAAAGCTTACAAATTTGAAAGAAAGAAAGCTTCT	840
DB	1318	AAGAGATGAGATGCTTTTATAGTAAAGCTTACAAATTTGAAAGAAAGAAAGCTTCT	1377
QY	841	ATATCTCCAACTTCAATTTTCTGGGCCAACTCTGACTATGAGAAAGATTAAAG	900
DB	1378	ATATCTCCAACTTCAATTTTCTGGGCCAACTCTGACTATGAGAAAGATTAAAG	1437
QY	901	CAGCTGAGCATCAGGCGCAAAAGCAAACTCAAGCTGTGCACTTGGAAGCCAAAT	960
DB	1438	CAGCTGAGCATCAGGCGCAAAAGCAAACTCAAGCTGTGCACTTGGAAGCCAAAT	1497
QY	961	GAACCTGTCCTGCTGTCTCAGAGGGTGAAGAAAGCGAGAGCCCTCACTCACTCC	1020
DB	1498	GAACCTGTCCTGCTGTCTCAGAGGGTGAAGAAAGCGAGAGCCCTCACTCACTCC	1557
QY	1021	TGTGCCACTCTGCTACTCTCAGAGGCAGAGCAAAAGCCCGTGCATCCCGCAGCGTG	1080
DB	1558	TGTGCCACTCTGCTACTCTCAGAGGCAGAGCAAAAGCCCGTGCATCCCGCAGCGTG	1617
QY	1081	CCAGCGTCCAGCGTGCACCGCTCGTGTAAAGACAGCCCGCTGTGTACAGCGCTC	1140
DB	1618	CCAGCGTCCAGCGTGCACCGCTCGTGTAAAGACAGCCCGCTGTGTACAGCGCTC	1677
QY	1141	AGTGGCTGCACTGTCCGACAGAGCTGGAAGACAAATAAGCTCAAGCCTTCCCTC	1200
DB	1678	AGTGGCTGCACTGTCCGACAGAGCTGGAAGACAAATAAGCTCAAGCCTTCCCTC	1737
QY	1201	TCTCTGATATCAATCAATGATTCTATATTACGACGATGCGAGCATCTTACATGCTTC	1260
DB	1738	TCTCTGATATCAATCAATGATTCTATATTACGACGATGCGAGCATCTTACATGCTTC	1797
QY	1261	TCTCATCAGAAAGTGTGGTGGAACTACCAAACTTCCACTCTGTGATGGGACCAAC	1320
DB	1798	TCTCATCAGAAAGTGTGGTGGAACTACCAAACTTCCACTCTGTGATGGGACCAAC	1857
QY	1321	AAGCTATGCAATTCTCCCTGTTCAAGAACTATCCGAGGAGACTCCGAGAAACAGTCT	1380
DB	1858	AAGCTATGCAATTCTCCCTGTTCAAGAACTATCCGAGGAGACTCCGAGAAACAGTCT	1917
QY	1381	GATTAAGAGAGAGCAGACATCCCAAGAGCTGACACCGCAGGCTTCAACAGCCAG	1440
DB	1918	GATTAAGAGAGAGCAGACATCCCAAGAGCTGACACCGCAGGCTTCAACAGCCAG	1977
QY	1441	AGCAAGGATTTGCAATTTGCTCAGAACCAAGAGGAGGCAACCGCCAGAGGTCCTTTTA	1500
DB	1978	AGCAAGGATTTGCAATTTGCTCAGAACCAAGAGGAGGCAACCGCCAGAGGTCCTTTTA	2037
QY	1501	TCTCAGCTGATGAAAGTGGAGCGTGAAGACCAATTTACCAACCAAGCTTCTTTGGG	1560
DB	2038	TCTCAGCTGATGAAAGTGGAGCGTGAAGACCAATTTACCAACCAAGCTTCTTTGGG	2097
QY	1561	CTTTCCACGACGACGACCTCAAGAACTGCTGCGCTGGGCTTAAAGGCTGGCAC	1620
DB	2098	CTTTCCACGACGACGACCTCAAGAACTGCTGCGCTGGGCTTAAAGGCTGGCAC	2157
QY	1621	TCCGATATCTTGGCCCGCCAGACCTCTACCCCTTCCCTGACAGAGCGTGTATTTGGC	1680
DB	2158	TCCGATATCTTGGCCCGCCAGACCTCTACCCCTTCCCTGACAGAGCGTGTATTTGGC	2217
QY	1681	ACGAGTCTCACAATTCTACTGTGCTCAGCATGACGAGGAGTGCAGTTACTCT	1740

[illegible]

LOCUS	AX482478	5450 bp	DNA	linear	PAT 16-AUG-2002
DEFINITION	Sequence 147 from Patent WO20257460.				
ACCESSION	AX482478				
VERSION	AX482478.1	GI:22316999			

SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

REFERENCE
AUTHORS
1
Toddard, C.G., Bol, D., Finger, J., Lee, L., Nelson, T., Schieven, G.,

TITLE Polynucleotides encoding human phosphatases
JOURNAL Patent: WO 02057460-A 147 25-JUL-2002;
SQUIBB BRISTOL MYERS CO (US)

FEATURES	Location/Qualifiers
source	1. .5450

```

/db_xref="taxon:9606"
538. .2535
CDS

```

```

/sequence_id="2"
/protein_id="CAD44453.1"
/db_xref="GI:22317000"
/db_xref="REMBL:CAD44453"
/translation="MAHEMIGQIVTEELVLLSGETKVLIDSRPVEVNTSHIEI
AININCSIMKMRRLQODKVLITELIHSKAKKVIDCSQKVVYDQSDPAVSLSDC
FLYVLGLTEKSRNSVHLIAGSFAEFSCFPGELCEGKSTLYPTCTISDCELVANIGCP
RILPMLDGVCDVNLKELIQONGIYGLVANSYTCPRKDFIPESHFLPDAVDFKCE
ILPMLDKSDVILERAKANGCVLHYICLANSITAIAYIMKRDMSILPNDVSVFVKE
KRPTISPMFNFLGOLIDYEKKIKNOTGASGPKSKILHLKEPMEPVAVSEGGKSC
TLSPGPCADSTSEAGORPVHPASVSPVQPSLESDLPVALSGHLSAPRLKE
SNKIKRPSRLDIKSYVSASMAALSHGSESDMLFYKKPSTTIDGNKILGSPVQDS
LSEQTPETSPDKESASTPKULQTAHPSPDSQSKRLHSVTSSSGTRQKRLSLPHRSQ
VEDNHTFTFLGLSTSQOHLTKSAGLITKGMHSDILAQTSPTSLTSSWYATFSSSH
YASASLYGSGASVSAYVSCSLPTCGDYVYVRRKRCQDPRADSRSMHEESPFEKQF
RRSCSMGEGSIMSENREBELGKGYGSGSSFGSNSEILIEVA"

```

ORIGIN

	Query Match Similarity	100.0%	Score 1995	DB 6	Length 5450
Beef Local	Similarity	100.0%	Pred. No. 0		
Matches 1995	Conservative	0	Mismatches	0	Indels 0
					Gaps 0
Qy	1	ATGGCCCATGAGATGATTGGAACTCAAAATGTGTACTGAGAGGTTGTGTCCTCTGGTGA	60		
Db	538	ATGGCCCATGAGATGATTGGAACTCAAAATGTGTACTGAGAGGTTGTGTCCTCTGGTGA	597		

Oy	61	AGTGAACGGAAAAAGTCCTGTAATGATAGCCGGCAATTTGGGATPACAATACATCC	120
Dp	558	AGTGAACGGAAAAAGTCCTGTAATGATAGCCGGCAATTTGGGAAATPACAATACATCC	657
Oy	121	CACATTTTGGAAAGCATTATATCACTGCTCCAAAGCTTATGAAACGAAGTTGGACAAG	180
Dp	658	CACATTTTGGAAAGCATTATATCACTGCTCCAAAGCTTATGAAACGAAGTTGGACAAG	717
Oy	181	GACAAAGTTATATTACAAGCTCATCAGCATTCAGCGAAACATTAAGTTGACATTGAT	240
Dp	718	GACAAAGTTATATTACAAGCTCATCAGCATTCAGCGAAACATTAAGTTGACATTGAT	777
Oy	241	TGCAGTCGAAGGTTGTAGTTACATCAAAAGCTCCCAAGATGTTGCCCTCTCTCTTCA	300
Dp	778	TGCAGTCGAAGGTTGTAGTTACATCAAAAGCTCCCAAGATGTTGCCCTCTCTCTTCA	837
Oy	301	GACTGTTTTCACGTGACTCTGGGGTAACTGGAAGAGCTTCAACTCTTGCACCTG	360
Dp	838	GACTGTTTTCACGTGACTCTGGGGTAACTGGAAGAGCTTCAACTCTTGCACCTG	897

QY	421	ACTGAGTCCCTACCTGGAATTTCTCAGGCTGTCTTACCTGTTGGCAACATTTGGGCCAAC	480
Db	968	ACTGAGTCCCTACCTGCAATTTCTCAGGCTGTCTTACCTGTTGGCAACATTTGGGCCAAC	101
QY	481	CGAATTCCTCCCAATCTTTATCTTGGCTGCACGCAAGATGTCTCTCAACAGGAGCTGATA	540
Db	1018	CGAATTCCTCCCAATCTTTATCTTGGCTGCACGCAAGATGTCTCTCAACAGGAGCTGATA	107
QY	541	CAGCGAATGGGATTTGGTTATGTGTAAATGCGACACTATCTCTGTCCAAAGCTGACTTT	600
Db	1078	CAGCGAATGGGATTTGGTTATGTGTAAATGCGACACTATCTCTGTCCAAAGCTGACTTT	113
QY	601	ATCCCGGAGTCTCATTTCTGCGGTGTGCTGTGAATGACAGCTTTTGTGAGAAAATTTTG	660
Db	1138	ATCCCGGAGTCTCATTTCTGCGGTGTGCTGTGAATGACAGCTTTTGTGAGAAAATTTTG	119
QY	661	CCGTGTTGACCAATCAGTAGATTTTCATTTGAGAAAAGCAAAAGCTCCAAATGATGTGTT	720
Db	1198	CCGTGTTGACCAATCAGTAGATTTTCATTTGAGAAAAGCAAAAGCTCCAAATGATGTGTT	125
QY	721	CTAATGCACTGTTTAGCTGGGATCTCCCGCTCCGCAACATGCGTATGGCTTACATCATG	780
Db	1258	CTAATGCACTGTTTAGCTGGGATCTCCCGCTCCGCAACATGCGTATGGCTTACATCATG	131
QY	781	AAGAGGATGGACATGCTCTTATGATGAATGAAGCTTACAGATTTGTGAAAGAAAAGACCTACT	840
Db	1318	AAGAGGATGGACATGCTCTTATGATGAATGAAGCTTACAGATTTGTGAAAGAAAAGACCTACT	137
QY	841	ATATCTCCAAATCTTCAATTTTCTGGGCCAACTCTGCACTATAGAGAAGATTTAAGAAC	900
Db	1378	ATATCTCCAAATCTTCAATTTTCTGGGCCAACTCTGCACTATAGAGAAGATTTAAGAAC	143
QY	901	CAGACTGGAGCATCAGGGCCAAAGAGCAAACTCAAGCTGTCTGCACTGTGAGAAAGCCAAAT	960
Db	1438	CAGACTGGAGCATCAGGGCCAAAGAGCAAACTCAAGCTGTCTGCACTGTGAGAAAGCCAAAT	149
QY	961	GAACTGTCCCGTCTGTCTCAGAGGGGTGAGACAGAAAACGAGACGCCCTCAGTCCACCC	1020
Db	1498	GAACTGTCCCGTCTGTCTCAGAGGGGTGAGACAGAAAACGAGACGCCCTCAGTCCACCC	155
QY	1021	TGTGCGCACTGTGTACTCTCAGAGGCAAGAGCAAAAGGCCGTGTCAATCCGCGACGCTG	1086
Db	1558	TGTGCGCACTGTGTACTCTCAGAGGCAAGAGCAAAAGGCCGTGTCAATCCGCGACGCTG	161
QY	1081	CCCAACGTGCCAGGCTGACGCGCTGTGTTAGAGGACAGCGCGCTGTATCAGGCGCTC	1144
Db	1618	CCCAACGTGCCAGGCTGACGCGCTGTGTTAGAGGACAGCGCGCTGTATCAGGCGCTC	167

QY 1141 AGTGGGCTGCACCTGTCCGACAGAGGCTGGAAGACAAATPAAGCTCAAGCTTCTTC 1200
| | | | |
| | | | |
Db 1678 AGTGGGCTGCACCTGTCCGACAGAGGCTGGAAGACAAATPAAGCTCAAGCTTCTTC 1737
| | | | |
QY 1201 TCTCTGGAATCAAAACAGTCTTCAATTCAGCAGAGATGGGAGCATCTTACATGGCTTC 1260
| | | | |
| | | | |
Db 1738 TCTCTGGAATCAAAACAGTCTTCAATTCAGCAGAGATGGGAGCATCTTACATGGCTTC 1797
| | | | |
QY 1261 TCTCTGGAATCAAAACAGTCTTCAATTCAGCAGAGATGGGAGCATCTTACATGGCTTC 1320
| | | | |
| | | | |
Db 1798 TCTCTGGAATCAAAACAGTCTTCAATTCAGCAGAGATGGGAGCATCTTACATGGCTTC 1857
| | | | |
QY 1321 AAGCTATGCCAGTCTTCTCCCTGTTCAGAACTATCGGAGCAGACTCCGAAACCACTCT 1380
| | | | |
| | | | |
Db 1858 AAGCTATGCCAGTCTTCTCCCTGTTCAGAACTATCGGAGCAGACTCCGAAACCACTCT 1917
| | | | |
QY 1381 GATTAAGAGGAGAGCCAGCATCCCGAAGAGCTGACAGCCGAGCTTCAGACAGCCAG 1440
| | | | |
| | | | |
Db 1918 GATTAAGAGGAGAGCCAGCATCCCGAAGAGCTGACAGCCGAGCTTCAGACAGCCAG 1977
| | | | |
QY 1441 AGCAAGCATTTGATTCGGTTCAGAACAGAGCAGAGTGGACCCGCCAGAGTCCCTTTTA 1500
| | | | |
| | | | |
Db 1978 AGCAAGCATTTGATTCGGTTCAGAACAGAGCAGAGTGGACCCGCCAGAGTCCCTTTTA 2037
| | | | |
QY 1501 TCTCCACTGCATCGAAGTGGAGACGTGGAGAGCAATTAACAACAGCTCTCTTTCCGGC 1560
| | | | |
| | | | |
Db 2038 TCTCCACTGCATCGAAGTGGAGACGTGGAGAGCAATTAACAACAGCTCTCTTTCCGGC 2097
| | | | |
QY 1561 CTTTCCACAGCCAGCAGCACTCAGAAAGTGTGCTGGCTGGGCTTAAAGGCTGGCAC 1620
| | | | |
| | | | |
Db 2098 CTTTCCACAGCCAGCAGCACTCAGAAAGTGTGCTGGCTGGGCTTAAAGGCTGGCAC 2157
| | | | |
QY 1621 TCGGATATCTTGGCCCCCAGACCTCTACCCCTTCCCTGACAGCAGCTGTATTTTGGC 1680
| | | | |
| | | | |
Db 2158 TCGGATATCTTGGCCCCCAGACCTCTACCCCTTCCCTGACAGCAGCTGTATTTTGGC 2217
| | | | |
QY 1681 ACAGAGCTCTCACAATCTCTACCTGCTCAGCCATCAGAGAGAGAGTGCAGTACTCT 1740
| | | | |
| | | | |
Db 2218 ACAGAGCTCTCACAATCTCTACCTGCTCAGCCATCAGAGAGAGAGTGCAGTACTCT 2277
| | | | |
QY 1741 GCCTACAGCTGCAAGCCAGCTGCCCCACTTGGCGAGACCAAGTCTATTTCTGTGCGCAGCGG 1800
| | | | |
| | | | |
Db 2278 GCCTACAGCTGCAAGCCAGCTGCCCCACTTGGCGAGACCAAGTCTATTTCTGTGCGCAGCGG 2337
| | | | |
QY 1801 CAGAGCCAGATGACAGAGCTGACTGCGCGCGAGAGCTGGGATGAAGAGAGCCCTTTTGA 1860
| | | | |
| | | | |
Db 2338 CAGAGCCAGATGACAGAGCTGACTGCGCGCGAGAGCTGGGATGAAGAGAGCCCTTTTGA 2397
| | | | |
QY 1861 AAGCAGTTTAAAGCAGAGCTGCAATATGAAATTGGAGAGAGCATCATGTCAAGAAC 1920
| | | | |
| | | | |
Db 2398 AAGCAGTTTAAAGCAGAGCTGCAATATGAAATTGGAGAGAGCATCATGTCAAGAAC 2457
| | | | |
QY 1921 AAGTCACGGGAGAGAGTGGGAAAAGTGGGAGTCACTAGCTTTTTCGGGAGCATGAA 1980
| | | | |
| | | | |
Db 2458 AAGTCACGGGAGAGAGTGGGAAAAGTGGGAGTCACTAGCTTTTTCGGGAGCATGAA 2517
| | | | |
QY 1981 ATCATTAGAGTCTTC 1995
| | | | |
| | | | |
Db 2518 ATCATTAGAGTCTTC 2532
| | | | |

RESULT 3
AX260342 1998 bp DNA linear PAT 26-OCT-2001
LOCUS
DEFINITION Sequence 3 from Patent WO0173059.
ACCESSION AX260342
VERSION AX260342.1 GI:16509305
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE
1

AUTHORS Meyers, R.A.
TITLE 38692 and 21117: dual specificity phosphatase molecules and uses
JOURNAL Patent: WO 0173059-A 3 04-OCT-2001;
Millennium Pharmaceuticals, Inc. (US)
FEATURES
source 1..1998
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 94.9%; Score 1893; DB 6; Length 1998;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGGCCCATGAGATGATTTGAACTCAATTGTTACTGAGAGCTTGGCTCTCTGTGAA 60
| | | | |
| | | | |
Db 1 ATGGCCCATGAGATGATTTGAACTCAATTGTTACTGAGAGCTTGGCTCTCTGTGAA 60
| | | | |
QY 61 AGTGAAGGAAAAAGTCTGCTAATGATAGCCGCGCATTTGGGAATACATATATCC 120
| | | | |
| | | | |
Db 61 AGTGAAGGAAAAAGTCTGCTAATGATAGCCGCGCATTTGGGAATACATATATCC 120
| | | | |
QY 121 CACATTTTGAAGCATTATATATCACTCTCAAGCTTATGAAGCGAAGTTGCAACAG 180
| | | | |
| | | | |
Db 121 CACATTTTGAAGCATTATATATCACTCTCAAGCTTATGAAGCGAAGTTGCAACAG 180
| | | | |
QY 181 GACAAAGTTTAAATTAACAGAGCTCATCCAGATTCAGGAAACATTAAGTTGACATGAT 240
| | | | |
| | | | |
Db 181 GACAAAGTTTAAATTAACAGAGCTCATCCAGATTCAGGAAACATTAAGTTGACATGAT 240
| | | | |
QY 241 TGCAGTCAGAAAGTGTGTGTTTAAAGATCAAGCTCCCAAGATGTGCTCTCTCTTCA 300
| | | | |
| | | | |
Db 241 TGCAGTCAGAAAGTGTGTGTTTAAAGATCAAGCTCCCAAGATGTGCTCTCTCTTCA 300
| | | | |
QY 301 GACTGTTTCTCAGTGAATCTTCTGGGTAACCTGGAAGAGCTTCACTCTGTTCACTG 360
| | | | |
| | | | |
Db 301 GACTGTTTCTCAGTGAATCTTCTGGGTAACCTGGAAGAGCTTCACTCTGTTCACTG 360
| | | | |
QY 361 CTTCGAGTGGGTTTGTCTGAGATCTCTGCTTTCCCTGGGCTCTGTGAAGGAAAAATCC 420
| | | | |
| | | | |
Db 361 CTTCGAGTGGGTTTGTCTGAGATCTCTGCTTTCCCTGGGCTCTGTGAAGGAAAAATCC 420
| | | | |
QY 421 ACTCTAGTCCCTACCTGATTTCTGAGCTTGTCTTACTGTGCGCAACATTTGGGCAAC 480
| | | | |
| | | | |
Db 421 ACTCTAGTCCCTACCTGATTTCTGAGCTTGTCTTACTGTGCGCAACATTTGGGCAAC 480
| | | | |
QY 481 CGAATTTCTCCCAATCTTATCTTGGCTGCGCAGAGATGCTCTCAACAAGAGCTGATA 540
| | | | |
| | | | |
Db 481 CGAATTTCTCCCAATCTTATCTTGGCTGCGCAGAGATGCTCTCTCAACAAGAGCTGATA 540
| | | | |
QY 541 CAGCAGATGGGATTTGCTTATCTGTTAAATGCACTTACCTGTCCAAAGCTGACTCT 600
| | | | |
| | | | |
Db 541 CAGCAGATGGGATTTGCTTATCTGTTAAATGCACTTACCTGTCCAAAGCTGACTCT 600
| | | | |
QY 601 ATCCCGAGTCTCATTTCTCTGCTGCTGCTGATGACAGCTTTTGTGAAGAAATTTTG 660
| | | | |
| | | | |
Db 601 ATCCCGAGTCTCATTTCTCTGCTGCTGCTGATGACAGCTTTTGTGAAGAAATTTTG 660
| | | | |
QY 661 CCGTGGTTGACAAATCAGTAGATTTCAATGGAAGAAAGAGCTCCAAATGGAATGTT 720
| | | | |
| | | | |
Db 661 CCGTGGTTGACAAATCAGTAGATTTCAATGGAAGAAAGAGCTCCAAATGGAATGTT 720
| | | | |
QY 721 CTAGTGCATGTTTATGCTGGGATCTCCGCTCCGACCATGCTATGCTCATCATCATG 780
| | | | |
| | | | |
Db 721 CTAGTGCATGTTTATGCTGGGATCTCCGCTCCGACCATGCTATGCTCATCATCATG 780
| | | | |
QY 781 AAGAGATGACATGCTTTAGATGAGCTTCAAGATTTGTGAAGAAAAAGACTTACT 840
| | | | |
| | | | |
Db 781 AAGAGATGACATGCTTTAGATGAGCTTCAAGATTTGTGAAGAAAAAGACTTACT 840
| | | | |
QY 841 ATATCTCCAACTTCAATTTTCTGGGCCCACTCTGAGCTATGAGAGAAAGTTAAGAAC 900
| | | | |
| | | | |

```

Db      841 ATATCTCAAAATTCAATTTCTGGGCAACTCTGAGACTATGAGAAGATTATGAAAC 900
Qy      901 CAGACTGAGACATCAGGGCCAAAGAGCAAACTCAAGCTGTGCACTCGAAGAACCAAT 960
Db      901 CAGACTGAGACATCAGGGCCAAAGAGCAAACTCAAGCTGTGCACTCGAAGAACCAAT 960
Qy      961 GAACCTGTCCCTGCTGCTCAGAGGGTGGACAGAAAAAGGAGAGCCCTCAGTCCACCC 1020
Db      961 GAACCTGTCCCTGCTGCTCAGAGGGTGGACAGAAAAAGGAGAGCCCTCAGTCCACCC 1020
Qy      1021 TGTGCGCACTGTGCTACTCAGAGGAGCAGAGCAAAAGCCGTGATCCCGCAGCGTG 1080
Db      1021 TGTGCGCACTGTGCTACTCAGAGGAGCAGAGCAAAAGCCGTGATCCCGCAGCGTG 1080
Qy      1081 CCCAGAGTCCCGCAGCGTGCAGCGCTGTGATGAGAGACGCCCTGTGATCAGGCGCTC 1140
Db      1081 CCCAGAGTCCCGCAGCGTGCAGCGCTGTGATGAGAGACGCCCTGTGATCAGGCGCTC 1140
Qy      1141 AGTGGGCTGCACCTGTCCGAGACAGGCTGGAAGACGAATAAGCTCAAGCGTCTTC 1200
Db      1141 AGTGGGCTGCACCTGTCCGAGACAGGCTGGAAGACGAATAAGCTCAAGCGTCTTC 1200
Qy      1201 TCTCTGGAATCAAAATCAGTTTATATTCAGCAGCATGCGACATCTTACATGAGCTTC 1260
Db      1201 TCTCTGGAATCAAAATCAGTTTATATTCAGCAGCATGCGACATCTTACATGAGCTTC 1260
Qy      1261 TCTCTATCAGAGATGCTTTGGAATACTAACAACTTCCACTATCTGTGATGGAGCAAC 1320
Db      1261 TCTCTATCAGAGATGCTTTGGAATACTAACAACTTCCACTATCTGTGATGGAGCAAC 1320
Qy      1321 AAGCTATGCAAGTTTCTCCCTGTTTCAAGAACTATCCGAGAGACTCCCGAAACCACTCT 1380
Db      1321 AAGCTATGCAAGTTTCTCCCTGTTTCAAGAACTATCCGAGAGACTCCCGAAACCACTCT 1380
Qy      1381 GATTAAGAGAGAACCCAGCATCCCAAGAAAGCTCAGACCGCCGCTTACAGACACCAAG 1440
Db      1381 GATTAAGAGAGAACCCAGCATCCCAAGAAAGCTCAGACCGCCGCTTACAGACACCAAG 1440
Qy      1441 AGCAGCGATTCGATTCGCTCAGAACCCAGCAGCAGGAGCCGCGCAAGAGTCCCTTTTA 1500
Db      1441 AGCAGCGATTCGATTCGCTCAGAACCCAGCAGCAGGAGCCGCGCAAGAGTCCCTTTTA 1500
Qy      1501 TCTCCACTGATGAGAGTGGAGCGTGGAGAGCAATTAACAACGAGTCTCTTTTCGGC 1560
Db      1501 TCTCCACTGATGAGAGTGGAGCGTGGAGAGCAATTAACAACGAGTCTCTTTTCGGC 1560
Qy      1561 CTTTCACAGCAGCAGCAGCACTCAGAAAGTCTGTGGCTTGAAGGCTTGAAGGCTGACAC 1620
Db      1561 CTTTCACAGCAGCAGCAGCACTCAGAAAGTCTGTGGCTTGAAGGCTTGAAGGCTGACAC 1620
Qy      1621 TCGGATATCTTGGCCCCCAGACCTTACCCCTTCCCTGACAGCAGCTGTGATTTTGGC 1680
Db      1621 TCGGATATCTTGGCCCCCAGACCTTACCCCTTCCCTGACAGCAGCTGTGATTTTGGC 1680
Qy      1681 ACGAGTCTCTCAACATTTTACTCTGCTCAGCCATCTAAGAGAGGCACTGCAATTTACTCT 1740
Db      1681 ACGAGTCTCTCAACATTTTACTCTGCTCAGCCATCTAAGAGAGGCACTGCAATTTACTCT 1740
Qy      1741 GCTTACAGCTGACAGCAGCTGCCACTTGGAGAGCAAAAGTCTAATCTGTGAGCAGAGCG 1800
Db      1741 GCTTACAGCTGACAGCAGCTGCCACTTGGAGAGCAAAAGTCTAATCTGTGAGCAGAGCG 1800
Qy      1801 CAGAGCCAGATGACAGAGCTGACTCGCGCGAGAGCTGGCATGAGAGAGAGCCCTTTTGA 1860
Db      1801 CAGAGCCAGATGACAGAGCTGACTCGCGCGAGAGCTGGCATGAGAGAGAGCCCTTTTGA 1860
Qy      1861 AAGCAGTTTAAAGCAGAGAGCTGCCAATGGAATTTGAGAGAGCATCATGTGAGGAAC 1920
Db      1861 AAGCAGTTTAAAGCAGAGAGCTGCCAATGGAATTTGAGAGAGCATCATGTGAGGAAC 1920
Qy      1921 AAGTCAAGGAGAGAGCTGGGGAAGTGGGAGTCAGCTTCTTCTGGGAGCAGATGGA 1980
Db      1921 AAGTCAAGGAGAGAGCTGGGGAAGTGGGAGTCAGCTTCTTCTGGGAGCAGATGGA 1980

```

```

Db      1921 AAGTCAAGGAGAGAGCTGGGGAAGTGGGAGTCAGCTTCTTCTGGGAGCAGATGGA 1980
Qy      1981 ATCATTGAGTCTCC 1995
Db      1981 ATCATTGAGTCTCC 1995

RESULT 4
AX278461
LOCUS      AX278461          3059 bp      DNA      linear      PAT 02-NOV-2001
DEFINITION Sequence 1 from Patent WO0177340.
ACCESSION  AX278461
VERSION     AX278461.1  GI:16605915
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini, Hominidae; Homo.

REFERENCE   1
  AUTHORS   Duecker, K.
  TITLE     Identification of a dual specificity phosphatase: dusp-10
  JOURNAL   Patent: WO 0177340-A 1 18-OCT-2001;
            MERCK PATENT GmbH (DE)

FEATURES
  source     1..3059
             /organism="Homo sapiens"
             /mol_type="unassigned DNA"
             /db_xref="taxon:9606"
             /note="unnamed protein product"
             /codon_start=1
             /protein_id="CAD10545.1"
             /db_xref="GI:1605916"
             /db_xref="EMBL:CAD10545"
             /translation="MAHEWITGQIYTERVALLESSTREKVLIDSRPPEVNTSHILE
             /translaltion="MAHEWITGQIYTERVALLESSTREKVLIDSRPPEVNTSHILE
             AININCSKLMKRLQODKVLITELI,OHSAKHVDIDSCQVAVVVOOSQDVASLSDDC
             PLTVLIGLEKSPFVHLLAGAFESRFPGLCGKSLVPTCSLOPCLPVANTGPT
             RLPLPLVIGCCQDVANKELMQONGCYLINAENTCPKPFIESHFLRPVNDPEK
             ILPLIDKSDFLIEKASNSGCVLHCLAGISRSATTAIYINKRDMDSIDEAYRFVKE
             KRPTLSIPNENFLGQILDYEKIKNGTGASGPKSKLHLLEKPNBPVAVSFGQKSE
             TPLSPPCADSSATSEBAAGQRPVAPASVPSPVPSLLEDSPLVOLA,SGHLASDRLED
             SKLRSFSLDLSKYSYSMAASLHGFSSSDALEYKPSYTLDTGNTLCOFSPVOS
             LSEOTPERTSPDKSEASIPKLOTARPDSQSRHSVRSSTGRTORSL,LSPLRHGGS
             VEDNHTSTLRELSOCHLTGSAGLGLKGMHSDILAPOTSPSLTSMYFPTRESHF
             YSAHLYGGSASYSAVSGQPLFCGQVYSVRKQRPSPDRADSRKSMHESPPEKQFK
             RRSQCMERGESIMSENREBELKVGSSFSFGSMELIEVS"

ORIGIN
Query Match      94.9%; Score 1893; DB 6; Length 3059;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy      1 ATGCCCATGATGATGATGGAATGGAATTTGTTACTGAGAGGTGGCTGCTGAGAA 60
Db      127 ATGCCCATGATGATGATGGAATGGAATTTGTTACTGAGAGGTGGCTGCTGAGAA 186
Qy      61 AGTGAACGGAAAAAGTGTCTTAATTGATAGCCCGGCATTTTGGAAATACATATCC 120
Db      187 AGTGAACGGAAAAAGTGTCTTAATTGATAGCCCGGCATTTTGGAAATACATATCC 246
Qy      121 CACATTTTGGAAAGCCATTATATCACTGCTCCAGCTTATGAAAGGAGTTGCAACAG 180
Db      247 CACATTTTGGAAAGCCATTATATCACTGCTCCAGCTTATGAAAGGAGTTGCAACAG 306
Qy      181 GACAAAGTTTAAATTAACAGAGCTCATCCAGATTTCAGGCAAACTAAGGTTACATTTGAT 240
Db      307 GACAAAGTTTAAATTAACAGAGCTCATCCAGATTTCAGGCAAACTAAGGTTACATTTGAT 366
Qy      241 TGCAGTCAGAAAGTTGTAGTTTACATCAACAGCTCCCAAGATGTTGCTCTCTTCA 300
Db      367 TGCAGTCAGAAAGTTGTAGTTTACATCAACAGCTCCCAAGATGTTGCTCTCTTCA 426
Qy      301 GACTGTTTCTACGTACTCTTCTGGGTAAACTGAGAGAGCTTCAACTCTGTTACCTG 360

```


Dd	427	GAAGTCTTTCTCACTGACTCTTGGGTAAAC	TGGAGAAAGCTTAACTCTGTTCACTG	486
Qy	361	CTTGCAGGATGGGTTTGTCTGAGTTCTCTG	TGTGTTCCCGGAGCTCTGTGAAGAAATCC	420
Dd	487	CTTGCAGGATGGGTTTGTCTGAGTTCTCTG	TGTGTTCCCGGAGCTCTGTGAAGAAATCC	546
Qy	421	ACTCTAGTCCCTAAGCTGATTTCTAGCCTT	GGCTTAACTGTTGGCAATTTGGGCTAAC	480
Dd	547	ACTCTAGTCCCTAAGCTGATTTCTAGCCTT	GGCTTAACTGTTGGCAATTTGGGCTAAC	606
Qy	481	CGAATTTCTCCCAATCTTTATCTTGGCTGC	GCAGCAGATGTCTCAACAAGAGCTGATA	540
Dd	607	CGAATTTCTCCCAATCTTTATCTTGGCTGC	GCAGCAGATGTCTCAACAAGAGCTGATA	666
Qy	541	CAGCAGAAATGGGATTTGGTTATGTGTTAA	ATGCGACGTATACCTGTCCAAAGCTGACTT	600
Dd	667	CAGCAGAAATGGGATTTGGTTATGTGTTAA	ATGCGACGTATACCTGTCCAAAGCTGACTT	726
Qy	601	ATCCCGGAGTCTCATTTCCGCGTGTCGTGA	TGATGACGCTTTTGAGAGAAATTTTG	660
Dd	727	ATCCCGGAGTCTCATTTCCGCGTGTCGTGA	TGATGACGCTTTTGAGAGAAATTTTG	786
Qy	661	CCGTGTTGACAAATCAGTAGATTTCAATTG	AGAAAGCAAAAGCTCCCAATGATGTGT	720
Dd	787	CCGTGTTGACAAATCAGTAGATTTCAATTG	AGAAAGCAAAAGCTCCCAATGATGTGT	846
Qy	721	CTAGTGCACTGTTTAGCTGGGATCTCCCG	CTCCGACATCCGATAGGCTCTACATCATG	780
Dd	847	CTAGTGCACTGTTTAGCTGGGATCTCCCG	CTCCGACATCCGATAGGCTCTACATCATG	906
Qy	781	AAGAGATGACATGCTCTTTAGATGACCTT	ACAGATTTGTGAAAGAAAGAAAGACCTAC	840
Dd	907	AAGAGATGACATGCTCTTTAGATGACCTT	ACAGATTTGTGAAAGAAAGAAAGACCTAC	966
Qy	841	ATATCTCCAAACTTCAATTTTCTGGGCCCA	ATCTCTGACATAGAGAAAGATTAAAGAC	900
Dd	967	ATATCTCCAAACTTCAATTTTCTGGGCCCA	ATCTCTGACATAGAGAAAGATTAAAGAC	1026
Qy	901	CAGACTGAGCATCAGGGCCAAAGAGCAAA	CTCAAGCTGTGCACTTGAGAGAGCCAAT	960
Dd	1027	CAGACTGAGCATCAGGGCCAAAGAGCAAA	CTCAAGCTGTGCACTTGAGAGAGCCAAT	1086
Qy	961	GAACCTGTCCCTGTCTGTCTCAGAGGGT	GACAAAGCGGACGCTCTCACTGTCACCC	1020
Dd	1087	GAACCTGTCCCTGTCTGTCTCAGAGGGT	GACAAAGCGGACGCTCTCACTGTCACCC	1146
Qy	1021	TGTGCGGACTCTGCTACCTCAGAGGACAG	AGAACAAAGCCCGTGTCAATCCCGCACGGTG	1080
Dd	1147	TGTGCGGACTCTGCTACCTCAGAGGACAG	AGAACAAAGCCCGTGTCAATCCCGCACGGTG	1206
Qy	1081	CCACAGCTGCCCCAGCGGTGACCGCTGT	TATAGAGACAGCCCGCTGTCAAGCGCTC	1140
Dd	1207	CCACAGCTGCCCCAGCGGTGACCGCTGT	TATAGAGACAGCCCGCTGTCAAGCGCTC	1266
Qy	1141	AGTGGGCTGCACTGTCCGCGACAGAGGT	GTGAAGACACCAATTAAGCTCAAGCTTCTTC	1200
Dd	1267	AGTGGGCTGCACTGTCCGCGACAGAGGT	GTGAAGACACCAATTAAGCTCAAGCTTCTTC	1336
Qy	1201	TCTCTGATATCAAACTCAGTTTCAATTT	CAGGCACATGTGGCGACATCTTCAATGTGCTC	1260
Dd	1327	TCTCTGATATCAAACTCAGTTTCAATTT	CAGGCACATGTGGCGACATCTTCAATGTGCTC	1386
Qy	1261	TCCTCATCAGAAAGATGCTTTGGAAAT	ACTACAACTTCACTACTCTGATGGGACCAAC	1320
Dd	1387	TCCTCATCAGAAAGATGCTTTGGAAAT	ACTACAACTTCACTACTCTGATGGGACCAAC	1446
Qy	1321	AAGCTATGCAAGTTTCTCCCTGTTCAGGA	ACTATGGAAGCACTTCCGAAACCACTGCTT	1380
Dd	1447	AAGCTATGCAAGTTTCTCCCTGTTCAGGA	ACTATGGAAGCACTTCCGAAACCACTGCTT	1506
Qy	1381	GATTAAGAGAAAGCCAGCATCCCCAAAG	ACTGCGACCGCTTTCAGACGACCG	1440

Db	1507	GATTAAGAGGAAAGCCAGACATCTCCCAAGAACTGCAGACCCCGAGCGCTTACAGACAGCAG	1566
Qy	1441	AGCAAGCGATTGCAATTCCGTCAAGAACCAAGACAGAGTGGCAACCGCCCAAGAGTCCCTTTTA	1500
Db	1567	AGCAAGCGAATTGCATTCCGTACAGAACCAAGACAGAGTGGCACCCGCCAGAGGTCCCTTTTA	1626
Qy	1501	TCTCCACTGCAATGGAATGGGAGAGGTGGAGAGCAATTACACACACAGCTTCCCTTTCGGC	1560
Db	1627	TCTCCACTGCAATGGAATGGGAGAGGTGGAGAGCAATTACACACACAGCTTCCCTTTCGGC	1686
Qy	1561	CTTTTCCAACAGCCAGCAGCACTTCAACGAGTCTGTGGCCTTGGGCTTTAAAGGGCTGGCAC	1620
Db	1687	CTTTTCCAACAGCCAGCAGCACTTCAACGAGTCTGTGGCCTTGGGCTTTAAAGGGCTGGCAC	1746
Qy	1621	TCCGATATCTTTGGGCCCCCAGACCTCTACCCCTTCCCTGACACAGACAGCTGTAATTTTGGC	1680
Db	1747	TCCGATATCTTTGGGCCCCCAGACCTCTACCCCTTCCCTGACACAGACAGCTGTAATTTTGGC	1806
Qy	1681	ACAGAGTCTCACTCTACTCTGAGCTCAGCCATCTACAGGAGCAGTGGCAGTACTCT	1740
Db	1807	ACAGAGTCTCTCACTCTACTCTGAGCTCAGCCATCTACAGGAGCAGTGGCAGTACTCT	1866
Qy	1741	GCTTACAGCTGCAGCCAGCTGGCCCATCTTGGCGAGACCAAGTCTATTCTGTGGCAGGCGG	1800
Db	1867	GCTTACAGCTGCAGCCAGCTGGCCCATCTTGGCGAGACCAAGTCTATTCTGTGGCAGGCGG	1926
Qy	1801	CAGAAAGCCAAAGTACAGAGCTGACTCGCGCGAGCTGGGATGAAGAGAGCCGCTTTGAA	1860
Db	1927	CAGAAAGCCAAAGTACAGAGCTGACTCGCGCGAGCTGGGATGAAGAGAGCCGCTTTGAA	1986
Qy	1861	AAGCAGTTTAAACGCAGACAGCTGCCTCAATATGAAATTTGGAGAGAGCATCATGTCAAGAAC	1920
Db	1987	AAGCAGTTTAAACGCAGACAGCTGCCTCAATATGAAATTTGGAGAGAGCATCATGTCAAGAAC	2046
Qy	1921	AGCTCAGCGGGAAGAGCTGGGGAAAGTGGGGAAGTCAAGTCACTTTTCGGGCAAGATGAA	1980
Db	2047	AGCTCAGCGGGAAGAGCTGGGGAAAGTGGGGAAGTCAAGTCACTTTTCGGGCAAGATGAA	2106
Qy	1981	ATCATTGAGGCTCTCC	1995
Db	2107	ATCATTGAGGCTCTCC	2121
RESULT 5			
AX441210			
LOCUS	AX441210	3496 bp	DNA
DEFINITION	Sequence 1 from Patent WO0226997.		linear
ACCESSION	AX441210		
VERSION	AX441210.1	GI:21665766	
KEYWORDS			
SOURCE			
ORGANISM			
	Homo sapiens (human)		
	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
REFERENCE			
AUTHORS	1		
TITLE	LucHE, R.M. and Wei, B.		
JOURNAL	Dsp-16 dual-specificity phosphatase		
	Patent: WO 0226997-A 1 04-APR-2002;		
	Ceptlyr, Inc. (US)		
FEATURES			
source	1..3496	Location/Qualifiers	
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
ORIGIN			
Query Match	94.9%;	Score 1893;	DB 6; Length 3496;
Best Local Similarity	99.9%;	Pred. No. 0;	
Matches 1993;	Conservative	0; Mismatches	2; Indels 0; Gaps 0;
Qy	1	ATGGCCCATGATGATGATGGAACCTCAATTTCTTACTGAGAGGTTGGCTCTGCTGGAA	60
Db	562	ATGGCCCATGATGATGATGGAACCTCAATTTCTTACTGAGAGGTTGGCTCTGCTGGAA	621

QY 61 AGTGAACGGAAGAGTGTCTAATGATAGCCGCGCATTTGTGGAATACATATACATCC 120
DB 622 AGTGAACGGAAGAGTGTCTAATGATAGCCGCGCATTTGTGGAATACATATACATCC 681
QY 121 CACATTTTGGAGCCATTAATCACTGCTCCAGCTTATGAGCGAGGTGGCAACAG 180
DB 682 CACATTTTGGAGCCATTAATCACTGCTCCAGCTTATGAGCGAGGTGGCAACAG 741
QY 181 GACAAAGTCTTAATACAGAGCTCATTCAGACATTCAGCGAAACATAAGGTGACATTTGAT 240
DB 742 GACAAAGTCTTAATACAGAGCTCATTCAGACATTCAGCGAAACATAAGGTGACATTTGAT 801
QY 241 TGCAGTCAGAAAGGTGATGATTCAGATCAAGATCCCAAGATGTTGCTCTCTCTCA 300
DB 802 TGCAGTCAGAAAGGTGATGATTCAGATCAAGATCCCAAGATGTTGCTCTCTCTCA 861
QY 301 GACTGTTTCTCATCTGACTTCTGGGTAACTGAGAGAGCTTCAACTCTGTTCACTG 360
DB 862 GACTGTTTCTCATCTGACTTCTGGGTAACTGAGAGAGCTTCAACTCTGTTCACTG 921
QY 361 CTTCAGAGTGGGTTTGTGAGTTCTCTGTTGTTTCCCTGGCTCTGTGAAGAAATCC 420
DB 922 CTTCAGAGTGGGTTTGTGAGTTCTCTGTTGTTTCCCTGGCTCTGTGAAGAAATCC 981
QY 421 ACTCTAGTCCCTCACTGATTCCTGAGCTTGTGCTTACCTGTTGCAACATTTGGGCCAAC 480
DB 982 ACTCTAGTCCCTCACTGATTCCTGAGCTTGTGCTTACCTGTTGCAACATTTGGGCCAAC 1041
QY 481 CGAATCTTCCCAATCTTATCTTGTGCTGAGCGAGAGATGCTTCAACAGAGCTGATA 540
DB 1042 CGAATCTTCCCAATCTTATCTTGTGCTGAGCGAGAGATGCTTCAACAGAGCTGATA 1101
QY 541 CAGCAGATGGGATGTTATGTTTAATGCCAGCTATACCTGTCCAAAGCTGACTTT 600
DB 1102 CAGCAGATGGGATGTTATGTTTAATGCCAGCTATACCTGTCCAAAGCTGACTTT 1161
QY 601 ATCCCGAGTCTATTTCTGCGGTGCTGAGATGAGCAGCTTTGTGAAATTTTG 660
DB 1162 ATCCCGAGTCTATTTCTGCGGTGCTGAGATGAGCAGCTTTGTGAAATTTTG 1221
QY 661 CCGTGTGTGCAAAATCAGTGAATTTCAATTAAGAAAGCAAAAGCTTCAATGATGTGT 720
DB 1222 CCGTGTGTGCAAAATCAGTGAATTTCAATTAAGAAAGCAAAAGCTTCAATGATGTGT 1281
QY 721 CTATGTGCACTGTTAGTGGGATCTCCGCTCCGCCCACTCGCTATCGCTACATCATG 780
DB 1282 CTATGTGCACTGTTAGTGGGATCTCCGCTCCGCCCACTCGCTATCGCTACATCATG 1341
QY 781 AAGAGGATGAGCAATGCTTATGATGAACTTACAGATTTGTGAAAGAAAGAAAGCTACT 840
DB 1342 AAGAGGATGAGCAATGCTTATGATGAACTTACAGATTTGTGAAAGAAAGAAAGCTACT 1401
QY 841 ATATCTCCAAACTTCAATTTTCTGGGCCCACTCTGAGCTATGAGAAAGATTTAAGAAC 900
DB 1402 ATATCTCCAAACTTCAATTTTCTGGGCCCACTCTGAGCTATGAGAAAGATTTAAGAAC 1461
QY 901 CAGACTGGAGCATCAGGCGCAAAAGAGCAAACTCAAGCTGTGCACTGTGAGAAAGCAAT 960
DB 1462 CAGACTGGAGCATCAGGCGCAAAAGAGCAAACTCAAGCTGTGCACTGTGAGAAAGCAAT 1521
QY 961 GAACTGTCTGCTGTCTGAGAGGTGAGCAGAAAGCGAGAGCGCCCTCAATCCACCC 1020
DB 1522 GAACTGTCTGCTGTCTGAGAGGTGAGCAGAAAGCGAGAGCGCCCTCAATCCACCC 1581
QY 1021 TGTGCGCACTCTGTACTCAGAGGCAAGCAAAAGCGCGTGCATCCCGCAGCGTG 1080
DB 1582 TGTGCGCACTCTGTACTCAGAGGCAAGCAAAAGCGCGTGCATCCCGCAGCGTG 1641
QY 1081 CCCAGCGTCCAGAGCTGAGCGCTGCTGTTAGAGGACAGCCCGCTGTATCAGCGCTC 1140
DB 1642 CCCAGCGTCCAGAGCTGAGCGCTGCTGTTAGAGGACAGCCCGCTGTATCAGCGCTC 1701

QY 1141 AGTGGCTGACCTGTCTCCGAGACAGGCTGGAAGACAGCAATTAAGCTCAAGCTTCTTC 1200
DB 1702 AGTGGCTGACCTGTCTCCGAGACAGGCTGGAAGACAGCAATTAAGCTCAAGCTTCTTC 1761
QY 1201 TCTCTGATATCAATCAGTTTCAATTTTCAAGCCAGATGAGCAGATCTTACATGCTTC 1260
DB 1762 TCTCTGATATCAATCAGTTTCAATTTTCAAGCCAGATGAGCAGATCTTACATGCTTC 1821
QY 1261 TCCCTATCAGAAAGTCTTGGAAATCTAACAACCTTCACTATCTGTGAGTGGACCAAC 1320
DB 1822 TCCCTATCAGAAAGTCTTGGAAATCTAACAACCTTCACTATCTGTGAGTGGACCAAC 1881
QY 1321 AAGCTATGCGAGTTCTCCCTGTTCCAGAACTATCCGAGCAGACTCCGAAACCACTGCT 1380
DB 1882 AAGCTATGCGAGTTCTCCCTGTTCCAGAACTATCCGAGCAGACTCCGAAACCACTGCT 1941
QY 1381 GATTAAGAGAGAGCCAGCATCTCCCAAGAGCTGACAGCCGACCTTACAGACCTCAG 1440
DB 1942 GATTAAGAGAGAGCCAGCATCTCCCAAGAGCTGACAGCCGACCTTACAGACCTCAG 2001
QY 1441 AGCAAGCATTCATTCGGTCCAGAAACACACAGAGTGGACAGCCCGAGAGTCCCTTTA 1500
DB 2002 AGCAAGCATTCATTCGGTCCAGAAACACACAGAGTGGACAGCCCGAGAGTCCCTTTA 2061
QY 1501 TCTCCACTGCATCGAAGTGGAGCGTGGAGAGCAATTAACACACAGCTTCTTTCGGC 1560
DB 2062 TCTCCACTGCATCGAAGTGGAGCGTGGAGAGCAATTAACACACAGCTTCTTTCGGC 2121
QY 1561 CTTTCCACAGCCAGCAGACACCTTACAGAACTGTGCTGGCTGGGCTTTAAGGGCTGGCAC 1620
DB 2122 CTTTCCACAGCCAGCAGACACCTTACAGAACTGTGCTGGCTGGGCTTTAAGGGCTGGCAC 2181
QY 1621 TCGGATATCTTGGCCCCCCCCAGACCTTACACCTTCCCTTACACAGCAGCTGTATTTTGGC 1680
DB 2182 TCGGATATCTTGGCCCCCCCCAGACCTTACACCTTCCCTTACACAGCAGCTGTATTTTGGC 2241
QY 1681 ACAGAGTCTTCAACATTTACTTGTGCTGAGCCATTAAGAGAGCAGTCCAGTACTCT 1740
DB 2242 ACAGAGTCTTCAACATTTACTTGTGCTGAGCCATTAAGAGAGCAGTCCAGTACTCT 2301
QY 1741 GCTTACAGTGCAGCCAGCTGCTGCCACTTGCAGAGACCAAGTCTATTTCTGTGCGCAGCGG 1800
DB 2302 GCTTACAGTGCAGCCAGCTGCTGCCACTTGCAGAGACCAAGTCTATTTCTGTGCGCAGCGG 2361
QY 1801 CAGAGCCCAAGTGAAGAGCTGACTGCGCGGAGAGCTGCGCATGAAAGAGCCCTTTGAA 1860
DB 2362 CAGAGCCCAAGTGAAGAGCTGACTGCGCGGAGAGCTGCGCATGAAAGAGCCCTTTGAA 2421
QY 1861 AAGCAGTTTAAACGAGAACTGCGCAATGGAATTTGAGAGAGCATGTGAGAGAAC 1920
DB 2422 AAGCAGTTTAAACGAGAACTGCGCAATGGAATTTGAGAGAGCATGTGAGAGAAC 2481
QY 1921 AGGTCAAGGGAAGAGCTGGGGAAGAGTGGGAGTCACTTACGTTTTCGGGCAAGATGAA 1980
DB 2482 AGGTCAAGGGAAGAGCTGGGGAAGAGTGGGAGTCACTTACGTTTTCGGGCAAGATGAA 2541
QY 1981 ATCATTGAGGTCTCC 1995
DB 2542 ATCATTGAGGTCTCC 2556

RESULT 6
AB052156
LOCUS AB052156 3521 bp mRNA linear PRI 18-OCT-2001
DEFINITION Homo sapiens MKP-7 mRNA for MAPK phosphatase-7, complete cds.
ACCESSION AB052156
VERSION AB052156.1 GI:13548676
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1

AUTHORS	Masuda, K., Shima, H., Watanabe, M. and Kikuchi, K.
TITLE	MKP-7, a novel mitogen-activated protein kinase phosphatase, functions as a shuttle protein
JOURNAL	J. Biol. Chem. 276 (42), 39002-39011 (2001)
MEDLINE	21486429
PUBMED	11489891
REFERENCE	2 (bases 1 to 3521)
AUTHORS	Masuda, K., Shima, H. and Kikuchi, K.
TITLE	Direct Submission
JOURNAL	Submitted (05-DEC-2000) Kouhei Masuda, Institute for Genetic

Submitted (05-DEC-2000) Kouhei Masuda, Institute for Genetic Medicine, Division of Biochemical Oncology and Immunology, Kita-Ku
kita5-jou nishi7-tyoume, Sapporo, Hokkaido 060-0815, Japan
(E-mail: kou@imm.hokudai.ac.jp, Tel: 81-11-706-5536,
Fax: 81-11-707-6839)

FEATURES	Location/Qualifiers
source	1. .3521

```

/organism="Homo sapiens"
/mol_type="RNA"
/db_xref="taxon:9606"
/chromosome="12"
/cell_line="Jurkat"
1..3521
/gene="MKP-7"
564..2561
CDS
7..2561

```

/tcaaa:aa:aa:"MAHEMIQIWIETHVALLBEGTEKRIILLISPPVENSINITE
 ANINOSKMKRRLQOHLMIETIOLSHAKHMLIDGOKVYVADDOSSALISJNC
 FLVILALKEKSNTHILLAGFAERFCPLCEGKSTIYPTICIQCP.VYANIGEP
 RILENPLVIGQORUKEMLOMCHGLAIKSNATIAIIPESHFAPLVONFCEK
 ILPMLDZGVEIRKAASNGCYVHGLASISATIAIIPESHFAPLVONFCEK
 KRTPISNENFLOQLIDYEKIKIKNQKSGSPKILHLKRPVPAVNEGOSKE
 TPLSPCEADTEALGQPVHAPVAPVOPSLIDLEKHPVPAVNEGOSKE
 SNLKRFSFLDISEVYSVSMASHLIFSSSDIAEYKPCITDGNKILCOFPEV
 LISQETPSPDEKBAIIPKLOLORPDSOSKSHVSTSSGTOARILSLPHRGS
 VEDNYHPSFLPGISQOCH.TSAGLGLKMHMSDLIAQTSIPBISVWYATISSH
 YSALIVGGASATSVASCOLPPLCGOVSIVARRKODSDRADRSRMBESPEKQK
 RSGQNEPFGSINSEKSPCEGQVSSVFGSGNEIIEVS"

ORIGIN

Query Match	94.9%	Score 1893	DB 9	Length 3521
Best Local Similarity	99.9%	Pred. No.	0	
Matches 1993	Conservative	0	Mismatches 2	Indels 0
				Gaps 0

QY 1 ATGGCCCATGAGATGATGGAACTCAAAATTGTTACTGAGAGAGTTGGCTGCTGCGAA 60

Db 564 ATGGCCCATGAGATGATGGAACTCAAAATTGTTACTGAGAGAGTTGGCTGCTGCGAA 6233

QY 61 AGTGAACCGGAAAAAGTGTCTGTAATTGATAGCCGCGCATTTGTGGAATACATATACATCC 120

Db 624 AGTGAACCGGAAAAAGTGTCTGCTGAATTGATTACCCGCGCATTTGTGGAATACATATACATCC 683

QY 121 CATATTTTGGAGACCATTAATATATCACTGCTCCAGAGCTTATGAAAGCGAAGGTTGCAACAG 180

Db 684 CATATTTTGGAGACCATTAATATATCACTGCTCCAGAGCTTATGAAAGCGAAGGTTGCAACAG 7433

QY 181 GACAAAGTGTAAATTACAGAGCTCATCCAGACTTCAGCGAAAATTAAGTGTGACATTGAT 240

Db 744 GACAAAGTGTAAATTACAGAGCTCATCCAGACTTCAGCGAAAATTAAGTGTGACATTGAT 8033

QY 241 TGCAGTCAGAAAGGTTGATGTTTACATCAACAACAACCTCCAGAAATGTTGGCCTCTCTTCA 300

Db 804 TGCAGTCAGAAAGGTTGATGTTTACATCAACAACAACCTCCAGAAATGTTGGCCTCTCTTCA 8633

QY 301 GACTGTTTTCTCACTGATCTTCCTGGGTAAATCGAGAAAGAGCTTCAACTCTGTCACTG 360

Db 864 GACTGTTTTCTCACTGATCTTCCTGGGTAAATCGAGAAAGAGCTTCAACTCTGTCACTG 9233

QY 361 CTTCGACAGTGGGTTTGCTGAGTTCTCTGTTGTTTCCCTGGGCTCTGTGAAAGAAAATCC 420

Db	921	CTTGCAAGTGGGTTTGCTGAGTTCTCTCGTTGTTTCCCTGAGCCCTCTGTGAAGAAATTC	983
OY	421	ACTCTAGTCCCTACCTGTCATTTCTCAGCCTTGCTTACCTGTGGCAACATTGGGCCAAC	480
Db	984	ACTCTAGTCCCTACCTGTCATTTCTCAGCCTTGCTTACCTGTGGCAACATTGGGCCAAC	1043
OY	481	CGAATTCCTCCCAATCTTTATCTTGGGTGCGACAGAGATGCTCTCAACAGAGCTGATA	540
Db	1044	CGAATTCCTCCCAATCTTTATCTTGGGTGCGACAGAGATGCTCTCAACAGAGCTGATG	1103
OY	541	CAGCAGATGAGGATTTGTTATGTGTTAAATGCCAGCTATACTGTGCCAAAGCTGACTTT	600
Db	1104	CAGCAGATGAGGATTTGTTATGTGTTAAATGCCAGCAATACCTGTCCAAAGCTGACTTT	1163
OY	601	ATCCCGAGTCTCATTTCTGCGGTGCTGCTGTGATGACAGCTTTTGTGAGAAATTTTG	660
Db	1164	ATCCCGAGTCTCATTTCTGCGGTGCTGCTGTGATGACAGCTTTTGTGAGAAATTTTG	1223
OY	661	CCGCGGTTTGAACAATGATGAGATTTCAATTGAGAAAGCAAAAGCTCAATGAGTGGT	720
Db	1224	CCGCGGTTTGAACAATGATGAGATTTCAATTGAGAAAGCAAAAGCTCAATGAGTGGT	1283
OY	721	CTAGTGCACGTGTTAGCTGGGATCTCCCGCTCGCCACCAATCGCTATCGCTACATCATG	780
Db	1284	CTAGTGCACGTGTTAGCTGGGATCTCCCGCTCGCCACCAATCGCTATCGCTACATCATG	1343
OY	781	AAGAGATGGAATGCTCTTATAGATGAGCTTACAGATTTGTGAAGAAAGAAAGCTACT	840
Db	1344	AAGAGATGGAATGCTCTTATAGATGAGCTTACAGATTTGTGAAGAAAGAAAGCTACT	1403
OY	841	ATATCTCCAAACTTCATTTTCTGGGGCAACTCTCGACATGATGAGAAAGATTAAAGC	900
Db	1404	ATATCTCCAAACTTCATTTTCTGGGGCAACTCTCGACATGATGAGAAAGATTAAAGC	1463
OY	901	CAGACTGAGGATCAGGGCCAAAGAGCAAACTCAAGCTGTGCACTTGAGAAAGCCAAT	960
Db	1464	CAGACTGAGGATCAGGGCCAAAGAGCAAACTCAAGCTGTGCACTTGAGAAAGCCAAT	1523
OY	961	GAACCTGTCCCTGTGCTCTCAGAGGGTGGACAGAAAAACGAGACGCCCTCAGTCCACCC	1020
Db	1524	GAACCTGTCCCTGTGCTCTCAGAGGGTGGACAGAAAAACGAGACGCCCTCAGTCCACCC	1583
OY	1021	TGTGCGCACTCTGCTACTCCTCAGAGGAGAGCAAGGACCCCGTGATCCCGCCAGAGGTG	1080
Db	1584	TGTGCGCACTCTGCTACTCCTCAGAGGAGAGCAAGGACCCCGTGATCCCGCCAGAGGTG	1643
OY	1081	CCCAAGCTGCCCAAGCTGACAGCCCTGCTGTTAGAGACAGCCCGCTGTGATCAGAGCTC	1140
Db	1644	CCCAAGCTGCCCAAGCTGACAGCCCTGCTGTTAGAGACAGCCCGCTGTGATCAGAGCTC	1703
OY	1141	AGTGGGCTGCACTGTCTCGCAGACAGCGCTGGAGACAGCAATTAAGCTCAGCTTCTTC	1200
Db	1704	AGTGGGCTGCACTGTCTCGCAGACAGCGCTGGAGACAGCAATTAAGCTCAGCTTCTTC	1763
OY	1201	TCTCTGATATCAAAATCAGTTTCAATTTACAGCCAGATGAGAGATCTTATCATAGGCTTC	1260
Db	1764	TCTCTGATATCAAAATCAGTTTCAATTTACAGCCAGATGAGAGATCTTATCATAGGCTTC	1823
OY	1261	TCCTCATCAGAAAGCTTTGGAATATACAAACCTTCCACTACTCTGTGATGGAGCCAAC	1320
Db	1824	TCCTCATCAGAAAGCTTTGGAATATACAAACCTTCCACTACTCTGTGATGGAGCCAAC	1883
OY	1321	AAGCTATGCCAGTTCTCCCTGTTCAGGAACTATCGAGAGACACTCCCGAAACCAATGCTCT	1380
Db	1884	AAGCTATGCCAGTTCTCCCTGTTCAGGAACTATCGAGAGACTCCCGAAACCAATGCTCT	1943
OY	1381	GATTAAGAGAGAGCAGACATCTCCCAAGAAGTGGAGACCCGAGGCTTACAGACGACAG	1440
Db	1944	GATTAAGAGAGAGCAGACATCTCCCAAGAAGTGGAGACCCGAGGCTTACAGACGACAG	2003
OY	1441	AGCAAGGATTTGCAATTCGGTCAGAAACAGCAGCAGTGGGACCGGCCAGAGGTCCCTTTTA	1500
Db	2004	AGCAAGGATTTGCAATTCGGTCAGAAACAGCAGCAGTGGGACCGGCCAGAGGTCCCTTTTA	2063

OY	1501	TCCTCACTGCATCGAAGTGGAGGGGTGGAGGACAATTACACACACAGCTTCCTTTGGGC	1560		
Db	2064	TCCTCACTGCATTCGAAGTGGAGGGGTGGAGGACATTACACACACAGCTTCCTTTGGGC	2123		
OY	1561	CTTTCCACACAGCCACGACACCTCAGCAAGTCTGTGGCCTTGGGCTTTAAGAGCGCTGGCAC	1620		
Db	2124	CTTTCCACACAGCCACGACACCTCAGCAAGTCTGTGGCCTTGGGCTTTAAGAGCGCTGGCAC	2183		
OY	1621	TCGGATATCTTTGGCCCCCAGAACCTCTACCCCTTCTCCAGACACAGCTGGTATTTTGGC	1680		
Db	2184	TCGGATATCTTTGGCCCCCAGAACCTCTACCCCTTCTCCAGACACAGCTGGTATTTTGGC	2243		
OY	1681	ACAGAGTCTCTACACATTTCTACTCTGCGCCCTACGCAATCTACGGAGGACGAGCGAGTTACTCT	1740		
Db	2244	ACAGAGTCTCTACACATTTCTACTCTGCGCCCTACGCAATCTACGGAGGACGAGCGAGTTACTCT	2303		
OY	1741	GCCCTACAGCTGCAAGCCAGCTGCGCCCACTTTCGGAGAGCCAAAGTCTATCTGTGCGAGGCGG	1800		
Db	2304	GCCCTACAGCTGCAAGCCAGCTGCGCCCACTTTCGGAGAGCCAAAGTCTATCTGTGCGAGGCGG	2363		
OY	1801	CAGAAAGCCAAAGTACACAGACTGACCTTCGCGGGGGAGCTGGCATGAAAGAGAGCCCTTTGAA	1860		
Db	2364	CAGAAAGCCAAAGTACACAGACTGACCTTCGCGGGGGAGCTGGCATGAAAGAGAGCCCTTTGAA	2423		
OY	1861	AAGCAGTTTAAACGAGAAAGCTGCGCAATGGAATTTGGAGAGAGCATATGTCAGAGAAC	1920		
Db	2424	AAGCAGTTTAAACGAGAAAGCTGCGCAATGGAATTTGGAGAGAGCATATGTCAGAGAAC	2483		
OY	1921	AGGTCACGCGGAGAGAGCTGGGGGAAAGTGGGCACTGACTAGCTTTTCGGGAGCAGATGGA	1980		
Db	2484	AGGTCACGCGGAGAGAGCTGGGGGAAAGTGGGCACTGACTAGCTTTTCGGGAGCAGATGGA	2543		
OY	1981	ATCATTTGAGGTCTCC	1995		
Db	2544	ATCATTTGAGGTCTCC	2558		
RESULT 7					
AX260340	AX260340	3544 bp	DNA	linear	PAT 26-OCT-2001
LOCUS	DEFINITION	Sequence 1 from Patent WO0173059.			
AX260340	ACCESSION	AX260340			
VERSION	AX260340.1	GI:16509303			
KEYWORDS					
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
REFERENCE	1				
AUTHORS	Meyers, R.A.				
TITLE	38692 and 21117: dual specificity phosphatase molecules and uses therefor				
JOURNAL	Patent: WO 0173059-A 1 04-OCT-2001;				
FEATURES	Millennium Pharmaceuticals, Inc. (US)				
Source	Location/Qualifiers				
	1..3544				
	/organism="Homo sapiens"				
	/mol_type="unassigned DNA"				
	/db_xref="taxon:9606"				
	589..2586				
	/codon="unamed protein product"				
	/note_start=1				
	/protein_id="CAD10218.1"				
	/db_xref="GI:16509304"				
	/catalation="MAHEMIGQIVTEHVALLSGTEKVLIDRSPEVEXYNTSHILE				
	AININCSKLMKRLQODKVLITELLOHSAKHKVIDCSOKVVVDDOSDQVVASLSPOC				
	FLVILVGLKESFNSVHLIAGFAFESCPGLCEKSTLIPTCISOPCLPVANIGPTP				
	RLIPVLVGCORPDLNKRIMQONGIYVLAISNTPCKRDLIPESHFLRPVNDSPCKE				
	LIPWLDKVDPIEKKASNGCVLVHCLAGISRSATIAIYIMKPMDSLDSEAYFVKE				
	RLPTSPNFPGLQDLYDEKKIKQGTGASGSKIKLHLKEKPEVPAVASEGQKSE				
	TPSPGDSFATSEBAQRPVPHVSPVSPVOPSLJEDSPVLQALSGHLASDRLDRE				
CDS					

ORIGIN

	Query Match	94.9%	Score 1893;	DB 6;	Length 3544;	
	Beet Local Similarity	99.9%	Pred. No. 0;			
	Matches 1993;	Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;
OY	1 ATGGCCCATGATGATTGGAACTCAAAATTGTACTAGAGAGTGGTGCGCTCTGCTGAAA	60				
Dd	589 ATGGCCCATGATGATTGGAACTCAAAATTGTACTAGAGAGTGGTGCGCTCTGCTGAAA	648				
OY	61 AGTGAACGGAAAAAGTCGCTTAATTGATAGCCGGCCATTGTGGAATAACAATACATCC	120				
Dd	649 AGTGAACGGAAAAAGTCGCTTAATTGATAGCCGGCCATTGTGGAATAACAATACATCC	708				
OY	121 CACATTTTGGAGGCCATTAAATATCAACTGCTCCAAGCTTATGAAGCGAAGTTGCACACG	180				
Dd	709 CACATTTTGGAGGCCATTAAATATCAACTGCTCCAAGCTTATGAAGCGAAGTTGCACACG	768				
OY	181 GACAAAGTGTAAATTAAGAAGCTATCAGAGCTATCAGAGAAACAATAAGSTTGAACATTGAT	240				
Dd	769 GACAAAGTGTAAATTAAGAAGCTATCAGAGCTATCAGAGAAACAATAAGSTTGAACATTGAT	828				
OY	241 TGCAGTCAGAAAGTTGTAGTTTACAGATCAAAAGCTCCCAAGATGTTGCTCTCTCTTCA	300				
Dd	829 TGCAGTCAGAAAGTTGTAGTTTACAGATCAAAAGCTCCCAAGATGTTGCTCTCTCTTCA	888				
OY	301 GACTGTTTTCTCACTGTACTCTTGCGGTAAACTGAGAAAGACTTCAACTCTGTTCACTG	360				
Dd	889 GACTGTTTTCTCACTGTACTCTTGCGGTAAACTGAGAAAGACTTCAACTCTGTTCACTG	948				
OY	361 CTTCGAGTGGGTGTGCTGAAGTTCTCTCGTTGTTTCCCTGCGCTCTGTGAAGAAATACC	420				
Dd	949 CTTCGAGTGGGTGTGCTGAAGTTCTCTCGTTGTTTCCCTGCGCTCTGTGAAGAAATACC	1008				
OY	421 ACTTAGTCCCTACCTGCATTTCTCACGCTTGCTTACTGTGCGCAACATTGGGCGCAACC	480				
Dd	1009 ACTTAGTCCCTACCTGCATTTCTCACGCTTGCTTACTGTGCGCAACATTGGGCGCAACC	1068				
OY	481 CGAATTTCTCCCAATCTTTATCTTGAGTCCACGAGATGTCCTCAACAAGAGCTGATA	540				
Dd	1069 CGAATTTCTCCCAATCTTTATCTTGAGTCCACGAGATGTCCTCAACAAGAGCTGATA	1128				
OY	541 CAGCAGAAATGGGATTTGGTTATGTGTAAATGCCCAGCTATACCTGTCCAAAGCCTCGACTTT	600				
Dd	1129 CAGCAGAAATGGGATTTGGTTATGTGTAAATGCCCAGCTATACCTGTCCAAAGCCTCGACTTT	1188				
OY	601 ATCCCGCAGTCAATTTCTGCGCGTGCCTGTGAATAACAAGCTTTTGTGAGAAATTTTG	660				
Dd	1189 ATCCCGCAGTCAATTTCTGCGCGTGCCTGTGAATAACAAGCTTTTGTGAGAAATTTTG	1248				
OY	661 CCGTGTGTGGAACAATCAGTAGATTTTATTGAGAAACAAGCCCTCAATGATGTGT	720				
Dd	1249 CCGTGTGTGGAACAATCAGTAGATTTTATTGAGAAACAAGCCCTCAATGATGTGT	1308				
OY	721 CTAGTGCATGTGTTAGCTGGGATCTCCCGCTCGGCGCAACATCGCTATCGCTTACATCATG	780				
Dd	1309 CTAGTGCATGTGTTAGCTGGGATCTCCCGCTCGGCGCAACATCGCTATCGCTTACATCATG	1368				
OY	781 AAGAGGATGGAACATGTCTTTAGAGAAAGCTTACAGATTTGTGAAAGAAAAAAGCCACT	840				
Dd	1369 AAGAGGATGGAACATGTCTTTAGAGAAAGCTTACAGATTTGTGAAAGAAAAAAGCCACT	1428				
OY	841 ATTATCTCCAAACTCAATTTTCTGGGCGCAACTCCCTGAGCTATGAGAAAGATTAAAGAC	900				
Dd	1429 ATTATCTCCAAACTCAATTTTCTGGGCGCAACTCCCTGAGCTATGAGAAAGATTAAAGAC	1488				
OY	901 CAGACTGAGAGATGAGGCCAAGAGCAAACTCAAGCTGCTGCACTTGAGAGACCAAT	960				

Db 1489 CAGACTGAGCATCAGGGCCAAAGACAACTCAAGCTGCTGCACCTGAGAAACCAAT 1548
Qy 961 GAACCTGTCCTGCTGCTCTCAGAGGGTGGACAGAAAAGCGAGCCCTCAGTCCACCC 1020
Db 1549 GAACCTGTCCTGCTGCTCTCAGAGGGTGGACAGAAAAGCGAGCCCTCAGTCCACCC 1608
Qy 1021 TGTGCGGACTGTGCTCCTCAGAGGACAGAGCAAAAGCCCTGCAATCCCGCAGGGTG 1080
Db 1609 TGTGCGGACTGTGCTCCTCAGAGGACAGAGCAAAAGCCCTGCAATCCCGCAGGGTG 1668
Qy 1081 CCCAGCGGCGCCAGCGTGCAGCCCTCGCTGTAAGAGCAGCCCGCTGGTAAGCGGCTC 1140
Db 1669 CCCAGCGGCGCCAGCGTGCAGCCCTCGCTGTAAGAGCAGCCCGCTGGTAAGCGGCTC 1728
Qy 1141 AGTGGGCTGCACCTGTCCGAGACAGGGTGGAGACAGCAATTAAGCTCAAGCCGTCCTTC 1200
Db 1729 AGTGGGCTGCACCTGTCCGAGACAGGGTGGAGACAGCAATTAAGCTCAAGCCGTCCTTC 1788
Qy 1201 TCTCTGAGATCAAAATCAGTTTCATATTCAGCCAGACAGAGCAGCATCTTACATGCTTC 1260
Db 1789 TCTCTGAGATCAAAATCAGTTTCATATTCAGCCAGACAGAGCAGCATCTTACATGCTTC 1848
Qy 1261 TCCCTCACAAGATGCTTGGTAATCTACAAACCTTCCACTACTCTGATGGAGACCAAC 1320
Db 1849 TCCCTCACAAGATGCTTGGTAATCTACAAACCTTCCACTACTCTGATGGAGACCAAC 1908
Qy 1321 AAGCTATGCCAGTCTCCCTGCTTCAAGAACTATCCGAGAGACCTCCCGAAACAGTCCCT 1380
Db 1909 AAGCTATGCCAGTCTCCCTGCTTCAAGAACTATCCGAGAGACCTCCCGAAACAGTCCCT 1968
Qy 1381 GATTAAGAGAGAGCGCAGCATCCCAAGAGCTGCAGACCGCCAGGCTTTCAGACACGAG 1440
Db 1969 GATTAAGAGAGAGCGCAGCATCCCAAGAGCTGCAGACCGCCAGGCTTTCAGACACGAG 2028
Qy 1441 AGCAAGGATTTGATTCGCTCAGAACCAAGAGAGAGTGGACCCCGCCAGAGTCCCTTTTA 1500
Db 2029 AGCAAGGATTTGATTCGCTCAGAACCAAGAGAGTGGACCCCGCCAGAGTCCCTTTTA 2088
Qy 1501 TCTCAGCTGATCAGAGTGGAGCGTGGAGAGCAATTAACCAACAGAGCTTCTTTTCGGC 1560
Db 2089 TCTCAGCTGATCAGAGTGGAGCGTGGAGAGCAATTAACCAACAGAGCTTCTTTTCGGC 2148
Qy 1561 CTTTCCACCGACGACGACCTTCAGAACTGCTGGCTGGCTTAAAGGCTTGGCAC 1620
Db 2149 CTTTCCACCGACGACGACCTTCAGAACTGCTGGCTGGCTTAAAGGCTTGGCAC 2208
Qy 1621 TCGGATATCTTTGGCCCCCGACCTCTAACCCTTCCCTGACCGACGAGCTGTATTTTGGC 1680
Db 2209 TCGGATATCTTTGGCCCCCGACCTCTAACCCTTCCCTGACCGACGAGCTGTATTTTGGC 2268
Qy 1681 ACAGAGTCTCAGACTTCTACTGTGCTCAGCCATCAGAGGAGGAGCTGCCAGTTACTCT 1740
Db 2269 ACAGAGTCTCAGACTTCTACTGTGCTCAGCCATCAGAGGAGGAGCTGCCAGTTACTCT 2328
Qy 1741 GCTTAAGCTGCGACGACGCTGCCACTTGGCGAGACCAAGTCTATTCTGTGCGCAGGCGG 1800
Db 2329 GCTTAAGCTGCGACGACGCTGCCACTTGGCGAGACCAAGTCTATTCTGTGCGCAGGCGG 2388
Qy 1801 CAGAAACCAAGTGCAGAGCTGACTCGCGCGGAGTGCATGAGAGAGAGGCCCTTTTGA 1860
Db 2389 CAGAAACCAAGTGCAGAGCTGACTCGCGCGGAGTGCATGAGAGAGAGGCCCTTTTGA 2448
Qy 1861 AAGCAGTTTAAAGCAGAAAGCTGCCAAATGGAATTTGAGAGAGCATCTGTCAAGAAAC 1920
Db 2449 AAGCAGTTTAAAGCAGAAAGCTGCCAAATGGAATTTGAGAGAGCATCTGTCAAGAAAC 2508
Qy 1921 AGGTCAACGGGAAGAGTGGGGAAGTGGGCACTCAGTCTAGCTTTTGGGCGACATGAA 1980
Db 2509 AGGTCAACGGGAAGAGTGGGGAAGTGGGCACTCAGTCTAGCTTTTGGGCGACATGAA 2568
Qy 1981 ATCATTTGAGTCTCC 1995
Db 2569 ATCATTTGAGTCTCC 2583

RESULT 8
AF506796
LOCUS
DEFINITION
AF506796 Homo sapiens dual specificity phosphatase 16 (DUSP16) mRNA,
complete cds.
ACCESSION
AF506796
VERSION
AF506796.1 GI:25573087
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 3566)
Hornaeft, I., Marynen, P., Goris, J., Sciote, R., and Baens, M.
MAPK phosphatase DUSP16/MKP-7, a candidate tumor suppressor for
chromosome region 12p12-13, reduces BCR-ABL-induced transformation
Oncogene 22 (49), 7728-7736 (2003)
JOURNAL
14586399
REFERENCE
2 (bases 1 to 3566)
Hornaeft, I., Marynen, P., and Baens, M.
Direct Submision
Submitted (26-Apr-2002) Department for Human Genetics-Flinders
Interuniversity Institute for Biotechnology (VIB), Katholieke
Universiteit Leuven, Herestraat 49, Leuven B-3000, Belgium
FEATURES
source
location/Qualifiers
1..3566
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="12"
/map="12p13-p12"
/tissue_type="bone marrow"
1..3566
/gene="DUSP16"
/note="synonym: MKP7"
633..2630
/gene="DUSP16"
/EC_number="3.1.3.48"
/EC_number="3.1.3.16"
/codon_start=1
/product="dual specificity phosphatase 16"
/protein_id="AA075120.1"
/db_xref="GI:25573088"

ORIGIN
Query Match 94.9%; Score 1893; DB 9; Length 3566;
Best local similarity 99.9%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 ATGGCCCATGAGATGATTGGAACCTCAATTTGTTACTGAGAGGTTGGTCTGCTGGAA 60
Db 633 ATGGCCCATGAGATGATTGGAACCTCAATTTGTTACTGAGAGGTTGGTCTGCTGGAA 692
Qy 61 AGTGAACGGAAGAGTGTCTAATGATAGCGGCACTTGTGGAATAACAATACATCC 120
Db 693 AGTGAACGGAAGAGTGTCTAATGATAGCGGCACTTGTGGAATAACAATACATCC 752
Qy 121 CACATTTGGAAGCATTAATCACTGCTCAAGCTTATGGAAGCAAGTTGCAACAG 180

Db 753 CACATTGGAAAGCCATTATATCACTGCTCCAGCTTATGAAAGGATGGCAACAG 812
Qy 181 GCAAGGTGTATTTACAGAGCTCATCCAGCATTCAGCGAAATATAGGTGACATGAT 240
Db 813 GACAAAGTGTATTAACAGAGCTCATCCAGCATTCAGCGAAATATAGGTGACATGAT 872
Qy 241 TGCAGTCAGAAAGGTGTAGTTTACATCAAAAGCTCCCAAGATGTTGCTCTCTCTCA 300
Db 873 TGCAGTCAGAAAGGTGTAGTTTACATCAAAAGCTCCCAAGATGTTGCTCTCTCTCA 932
Qy 301 GACTGTTTTCTCATCTGACTCTGCGGTAACTGGAAGAAGCTTCAACTCTGTTCACTG 360
Db 933 GACTGTTTTCTCATCTGACTCTGCGGTAACTGGAAGAAGCTTCAACTCTGTTCACTG 992
Qy 361 CTTCGAGGTGGGTTTCTGAGTCTCTGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
Db 993 CTTCGAGGTGGGTTTCTGAGTCTCTGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1052
Qy 421 ACTCTAGTCCCTACCTGACATTTCTCAGACCTTGTCTTACCTGTTGCAACATGGGCGCAC 480
Db 1053 ACTCTAGTCCCTACCTGACATTTCTCAGACCTTGTCTTACCTGTTGCAACATGGGCGCAC 1112
Qy 481 CGAATTCCTCCCAATCTTATATCTTGGCTGCGCAGAGATGCTCTCAACAGAGCTGATA 540
Db 1113 CGAATTCCTCCCAATCTTATATCTTGGCTGCGCAGAGATGCTCTCAACAGAGCTGATA 1172
Qy 541 CAGCAGAAATGGGATTTGTTATGTTTAAATGCGACGTAATACCTGTCGAAGCTGACTT 600
Db 1173 CAGCAGAAATGGGATTTGTTATGTTTAAATGCGACGTAATACCTGTCGAAGCTGACTT 1232
Qy 601 ATCCCGAGTCTCATTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660
Db 1233 ATCCCGAGTCTCATTTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1292
Qy 661 CCGTGTGTGACAAATCAGTATGATTTCAATGAGAAAGCAAAAGCTTCAATGATGATG 720
Db 1293 CCGTGTGTGACAAATCAGTATGATTTCAATGAGAAAGCAAAAGCTTCAATGATGATG 1352
Qy 721 CTATGTCACCTGTTTACCTGAGATCTCCGCTCCGCAACCTGCTTATGCTTACATGAT 780
Db 1353 CTATGTCACCTGTTTACCTGAGATCTCCGCTCCGCAACCTGCTTATGCTTACATGAT 1412
Qy 781 AAGAGATGGAACATGCTTTAGATGAAGCTTACAGATTTGTGAAGAAAGAAAGCACTACT 840
Db 1413 AAGAGATGGAACATGCTTTAGATGAAGCTTACAGATTTGTGAAGAAAGAAAGCACTACT 1472
Qy 841 ATATCTCCAAACTTCAATTTTCTGGGCAACTCTGACTATGAGAAGAAATTAAGAAC 900
Db 1473 ATATCTCCAAACTTCAATTTTCTGGGCAACTCTGACTATGAGAAGAAATTAAGAAC 1532
Qy 901 CAGACTGAGACATCAGGGCCAAAGAGCAAACTCAGCTGCTGCACTGGAAGAACCAAT 960
Db 1533 CAGACTGAGACATCAGGGCCAAAGAGCAAACTCAGCTGCTGCACTGGAAGAACCAAT 1592
Qy 961 GAACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1020
Db 1593 GAACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1652
Qy 1021 TGTGCGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
Db 1653 TGTGCGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1712
Qy 1081 CCCAGCGTGGCCAGCGCTGACGCGTGTGTATGAGAGACGCCGCTGCTGCTGCTGCTGCTG 1140
Db 1713 CCCAGCGTGGCCAGCGCTGACGCGTGTGTATGAGAGACGCCGCTGCTGCTGCTGCTGCTG 1772
Qy 1141 AGTGGGCTGACCTGTCGCGCAGACAGGCTGGAAGACAGCAATTAAGCTCAAGGCTCTCT 1200
Db 1773 AGTGGGCTGACCTGTCGCGCAGACAGGCTGGAAGACAGCAATTAAGCTCAAGGCTCTCT 1832
Qy 1201 TCTCTGATATCAAAATCAGTTTATATTCAGCCAGATGAGCAGATCTTACATGCTTC 1260
Db 1833 TCTCTGATATCAAAATCAGTTTATATTCAGCCAGATGAGCAGATCTTACATGCTTC 1892

Qy 1261 TCTCTATCAGAGATGCTTTGGAAATACACAAACCTTCCACTACTGTGATGGACCAAC 1320
Db 1893 TCTCTATCAGAGATGCTTTGGAAATACACAAACCTTCCACTACTGTGATGGACCAAC 1952
Qy 1321 AAGCTATGCGAGTTCTCCCTGTTTCCAGAACTATGAGACAGACTCCGAAACCACTGCT 1380
Db 1953 AAGCTATGCGAGTTCTCCCTGTTTCCAGAACTATGAGACAGACTCCGAAACCACTGCT 2012
Qy 1381 GATTAAGAGAAAGCCAGCATTCCTCCAAAGAGCTGACAGCCGCTTCAAGACCCAG 1440
Db 2013 GATTAAGAGAAAGCCAGCATTCCTCCAAAGAGCTGACAGCCGCTTCAAGACCCAG 2072
Qy 1441 AGCAGCGATTCATTCGCTGCAAGAACCCAGCAGAGTGGACCGCCAGAGTCCCTTTA 1500
Db 2073 AGCAGCGATTCATTCGCTGCAAGAACCCAGCAGAGTGGACCGCCAGAGTCCCTTTA 2132
Qy 1501 TCTCCACTGCACTGAAAGTGGAGCGTGGAGCAATTAACACACAGCTTCTTTTCCG 1560
Db 2133 TCTCCACTGCACTGAAAGTGGAGCGTGGAGCAATTAACACACAGCTTCTTTTCCG 2192
Qy 1561 CTTTCACAGCAGCAGACGACCTTACAGAAAGTCTGCTGCTGCTGCTTAAAGGCTG 1620
Db 2193 CTTTCACAGCAGCAGACGACCTTACAGAAAGTCTGCTGCTGCTGCTTAAAGGCTG 2252
Qy 1621 TGGGATATCTTGGCCCCCGAGACCTTACCCCTCCCTGACAGACAGCTGATTTTGC 1680
Db 2253 TGGGATATCTTGGCCCCCGAGACCTTACCCCTCCCTGACAGACAGCTGATTTTGC 2312
Qy 1681 ACAGAGTCTTACACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740
Db 2313 ACAGAGTCTTACACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2372
Qy 1741 GCTTACAGCTGACGACGCTGCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1800
Db 2373 GCTTACAGCTGACGACGCTGCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2432
Qy 1801 CAGAGCCAGTGAACAGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1860
Db 2433 CAGAGCCAGTGAACAGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2492
Qy 1861 AAGCAGTTTAAAGCAGAGCTGCCAAATGATTTGAGAGAGCATCATGTCAGAGAAC 1920
Db 2493 AAGCAGTTTAAAGCAGAGCTGCCAAATGATTTGAGAGAGCATCATGTCAGAGAAC 2552
Qy 1921 AGGTCAAGGAAAGAGCTGGGAAAGTGGGAGTCACTTACGCTTTTGGGCAAGATGAA 1980
Db 2553 AGGTCAAGGAAAGAGCTGGGAAAGTGGGAGTCACTTACGCTTTTGGGCAAGATGAA 2612
Qy 1981 ATCATTGAGGCTTCC 1995
Db 2613 ATCATTGAGGCTTCC 2627

RESULT 9
AX374994
LOCUS AX374994 3766 bp DNA linear PAT 01-MAR-2002
DEFINITION Sequence 17 from Patent WO0210363.
ACCESSION AX374994
VERSION AX374994.1 GI:19169826
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
AUTHORS Tang,Y.T., Elliott,V.S., Ramkumar,J., Yao,M.G., Buford,N.,
Wang,Y.E., Stewart,E.A., Gandhi,A.R., Peterson,C., Lee,E.A.,
Hatala,A.J., Lu,D.A., Tribouley,C.M., Griffin,J.A., Baughn,M.R.,
Yue,H., Warren,B.A., Nguyen,D.B. and Walla,N.K.
TITLE Protein phosphatases
Patent: WO 0210363-A 17 07-FEB-2002;
JOURNAL Incyte Genomics, Inc. (US)

FEATURES		Location/Qualifiers	
source		1..3766	/organism="Homo sapiens"
			/mol_type="unassigned DNA"
			/db_xref="taxon:9606"
ORIGIN		/note="Incycle ID No: 7480570CB1"	
Query Match 94.9%; Score 1893; DB 6; Length 3766;			
Best Local Similarity 99.9%; Pred. No. 0;			
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;			
Qy	1	ATGGCCCATGATGATTTGAACTCAAAATGTTACTGAGAGTGGTGGCTCTGCTGAA	60
Db	538	ATGGCCCATGAGATTTGAACTCAAAATGTTACTGAGAGTGGTGGCTCTGCTGAA	597
Qy	61	AGTGAACGGAAAAAGTCTGCTAATTGATAGCCGGCCATTGTTGGAAATACATATCC	120
Db	598	AGTGAACGGAAAAAGTCTGCTAATTGATAGCCGGCCATTGTTGGAAATACATATCC	657
Qy	121	CACATTTTGAAGCCATTAAATCAACTGCTCCAAGCTTATGAAGCGAAGTTGCCAAG	180
Db	658	CACATTTTGAAGCCATTAAATCAACTGCTCCAAGCTTATGAAGCGAAGTTGCCAAG	717
Qy	181	GACAAAGTTAATTACAGAGCTCATCCAGATTCAAGCAAAACATAAGTTGACATTGAT	240
Db	718	GACAAAGTTAATTACAGAGCTCATCCAGATTCAAGCAAAACATAAGTTGACATTGAT	777
Qy	241	TGCAGTCAGAAAGTTGATTGATTAGATCAAAAGCTCCCAAGATGTGGCTCTCTCTCA	300
Db	778	TGCAGTCAGAAAGTTGATTGATTAGATCAAAAGCTCCCAAGATGTGGCTCTCTCTCA	837
Qy	301	GACTGTTTCTCACTGTACTTCTGGGTAACTGAGAAAGACTTCAACTCTGTCACTG	360
Db	838	GACTGTTTCTCACTGTACTTCTGGGTAACTGAGAAAGACTTCAACTCTGTCACTG	897
Qy	361	CTTCGAGTGGGTTTGCAGAGTTCTCTCGTTGTTTCCCTGGCTCTGTGAAGAAATCC	420
Db	898	CTTCGAGTGGGTTTGCAGAGTTCTCTCGTTGTTTCCCTGGCTCTGTGAAGAAATCC	957
Qy	421	ACTCTAGTCCCTACCTGCAATTTCTCAGCCTTGTACTGTTGCTCAACATTTGGGCCAAC	480
Db	958	ACTCTAGTCCCTACCTGCAATTTCTCAGCCTTGTACTGTTGCTCAACATTTGGGCCAAC	1017
Qy	481	CGAATTTCTCCAAATCTTTATCTTGGCTGCCAGCGAGATGTCCTCAACAGAGCTGATA	540
Db	1018	CGAATTTCTCCAAATCTTTATCTTGGCTGCCAGCGAGATGTCCTCAACAGAGCTGATA	1077
Qy	541	CAGCAGAAATGGGATGGTATGTTGTTAAATGCCAGCTATACCTGTCCAAAGCTGACTTT	600
Db	1078	CAGCAGAAATGGGATGGTATGTTGTTAAATGCCAGCTATACCTGTCCAAAGCTGACTTT	1137
Qy	601	ATCCCGAGTCTCATTTTCTGCGTGTGCTGTGTAATGACAGCTTTTGTGAGAAAATTTTG	660
Db	1138	ATCCCGAGTCTCATTTTCTGCGTGTGCTGTGTAATGACAGCTTTTGTGAGAAAATTTTG	1197
Qy	661	CCGTGTTTGGACAAATCAGTAGATTTCATTGAGAAAGCAAAGCCTCCATGATGTGTT	720
Db	1198	CCGTGTTTGGACAAATCAGTAGATTTCATTGAGAAAGCAAAGCCTCCATGATGTGTT	1257
Qy	721	CTAGTCACTGTTTACCTGGGATCTCCGCTCCGACCATGCTATGCGCTCAATCATC	780
Db	1258	CTAGTCACTGTTTACCTGGGATCTCCGCTCCGACCATGCTATGCGCTCAATCATC	1317
Qy	781	AAGAGGATGAGCATGCTCTTTAGTAGAGTTTACAGATTTGTGAGAAAAAGACCTTACT	840
Db	1318	AAGAGGATGAGCATGCTCTTTAGTAGAGTTTACAGATTTGTGAGAAAAAGACCTTACT	1377
Qy	841	ATATCTCCAAACTTCAATTTTCTGGGCCCACTCCGAGCTATGAGAAAGATTTAGAAC	900
Db	1378	ATATCTCCAAACTTCAATTTTCTGGGCCCACTCCGAGCTATGAGAAAGATTTAGAAC	1437
Qy	901	CAGACTGGAGATCAGGGCCAAAGCAAACCTCAAGCTGTGCACTTGGAGAAAGCAAT	960

Db	1438	CAGACTGGAGCATCAGGGCCAAAGCAAACCTCAAGCTGTGCACTTGGAGAACCAAT	1497
Qy	961	GAACCTGTCCCTGCTCTCAGAGGGTGAACGAAAAGGAGAACGCCCTCAGTCCACCC	1020
Db	1498	GAACCTGTCCCTGCTCTCAGAGGGTGAACGAAAAGGAGAACGCCCTCAGTCCACCC	1557
Qy	1021	TGTGCCGACTCTGTCTAAGTTCAGAGGACGAGGACAAAGGCCGTGATCCGCCAGCGTG	1080
Db	1558	TGTGCCGACTCTGTCTAAGTTCAGAGGACGAGGACAAAGGCCGTGATCCGCCAGCGTG	1617
Qy	1081	CCGAGGTCGCCAGGGTGAAGCCGTGCTGTTAGAGACAGCCCGCTGTACAGGCGCTC	1140
Db	1618	CCGAGGTCGCCAGGGTGAAGCCGTGCTGTTAGAGACAGCCCGCTGTACAGGCGCTC	1677
Qy	1141	AGTGGCTGACACTGTCCGACAGAGGCTGGAAGACAAATTAAGCTCAAGGTTCCCTC	1200
Db	1678	AGTGGCTGACACTGTCCGACAGAGGCTGGAAGACAAATTAAGCTCAAGGTTCCCTC	1737
Qy	1201	TCTCTGATATCAAAATCAGTTTCATATTCAAGCAGATGGAGCATCTTACATGGCTTC	1260
Db	1738	TCTCTGATATCAAAATCAGTTTCATATTCAAGCAGATGGAGCATCTTACATGGCTTC	1797
Qy	1261	TCCTCATGAGAAAGATGCTTTGGAATACTAACAACCTTCCACTTCTGTGATGGGACCAAC	1320
Db	1798	TCCTCATGAGAAAGATGCTTTGGAATACTAACAACCTTCCACTTCTGTGATGGGACCAAC	1857
Qy	1321	AAGCTATCCAGTTCTCCCTGTTCAAGGACATATCGAGACGACTCCGAAACCAAGTCT	1380
Db	1858	AAGCTATCCAGTTCTCCCTGTTCAAGGACATATCGAGACGACTCCGAAACCAAGTCT	1917
Qy	1381	GATTAAGAGGAAAGCCAGATCCCAAGAGCTGCAGACCCGACGCTTACAGACGCAAG	1440
Db	1918	GATTAAGAGGAAAGCCAGATCCCAAGAGCTGCAGACCCGACGCTTACAGACGCAAG	1977
Qy	1441	AGCAAGCATTTGCAATTCGATCAGAACAGCAGAGTGGCAACCGCCAGAGGTCCTTTTA	1500
Db	1978	AGCAAGCATTTGCAATTCGATCAGAACAGCAGAGTGGCAACCGCCAGAGGTCCTTTTA	2037
Qy	1501	TCTCATCTGCATCGAAGTGGGAGCGTGGAGGACAAATTAACCAACCAAGCTTCTTTTGGC	1560
Db	2038	TCTCATCTGCATCGAAGTGGGAGCGTGGAGGACAAATTAACCAACCAAGCTTCTTTTGGC	2097
Qy	1561	CTTTCACAGCAGCAGAGCACTTCAGAGATCTGCGCTGGGCTTAAAGGCTGGCAC	1620
Db	2098	CTTTCACAGCAGCAGAGCACTTCAGAGATCTGCGCTGGGCTTAAAGGCTGGCAC	2157
Qy	1621	TGGATATCTTGGCCCCCAAGCCTTAACCTTCCCTGACAGCAGCTGTGATTTTGGC	1680
Db	2158	TGGATATCTTGGCCCCCAAGCCTTAACCTTCCCTGACAGCAGCTGTGATTTTGGC	2217
Qy	1681	ACAGATCTCTCAACTTTTCTCTGCTCAGGCACTTAAGGAGGAGTCCAGTTACTCT	1740
Db	2218	ACAGATCTCTCAACTTTTCTCTGCTCAGGCACTTAAGGAGGAGTCCAGTTACTCT	2277
Qy	1741	GCTTACAGCTGACAGCAGCTGCCACTTGGGAGGACCAAGTCTATTCTGTGGGCGGG	1800
Db	2278	GCTTACAGCTGACAGCAGCTGCCACTTGGGAGGACCAAGTCTATTCTGTGGGCGGG	2337
Qy	1801	CAGAGCCAAAGTACAGAGCTGACTGCGCGGAGAGCTGGCATGAAGAGAGCCCTTTGAA	1860
Db	2338	CAGAGCCAAAGTACAGAGCTGACTGCGCGGAGAGCTGGCATGAAGAGAGCCCTTTGAA	2397
Qy	1861	AAGCAGTTTAAAGCAGAGAGCTGCCAAATGAATTTGGAGAGACATCTGTCAAGAAC	1920
Db	2398	AAGCAGTTTAAAGCAGAGAGCTGCCAAATGAATTTGGAGAGACATCTGTCAAGAAC	2457
Qy	1921	AGGTCAACGGAAAGCTGGGGAAGTGGGCACTAGTCTTCTTGGGCGAGATGGA	1980
Db	2458	AGGTCAACGGAAAGCTGGGGAAGTGGGCACTAGTCTTCTTGGGCGAGATGGA	2517
Qy	1981	ATCATTAAGTCTCC 1995	

Db	1744	CTTTCACCAAGCAGACAGACCTTCAAGAACTCTGGCTTGGGCTTAAAGGCTTGGAC	1803
QY	1621	TCGGATATCTTTGGCCCCCAGACCTCTACCCCTTCCTGACCAAGAGAGCTGGATTTTGGC	1688
Db	1804	TCGGATATCTTTGGCCCCCAGACCTCTTACCCCTTCCCTGACCAAGAGAGCTGGATTTTGGC	1867
QY	1681	ACAGAATCTCTCACTTCTTACTCTTGGCTTACGCCATCTACGAGAGAGTGGCCAGTTACTCT	1740
Db	1864	ACAGAATCTCTCACTTCTTACTCTTGGCTTACGCCATCTACGAGAGAGTGGCCAGTTACTCT	1922
QY	1741	GCTTACAGCTGCAGCCAGCTGGCCCACTTGGCGGAGCAAGATCTATTCTGTGTGGCAGGCGG	1800
Db	1924	GCTTACAGCTGCAGCCAGCTGGCCCACTTGGCGGAGCAAGATCTATTCTGTGTGGCAGGCGG	1983
QY	1801	CAGAAAGCAATGAGCAGAGCTGACCTGGGGCGGAGACTGGCATGAGAGAGAGCCCCCTTTGAA	1867
Db	1984	CAGAAAGCAATGAGCAGAGCTGACCTGGGGCGGAGACTGGCATGAGAGAGAGCCCCCTTTGAA	2043
QY	1861	AAGCAGTTTAAACGCGAAGAGCTGCCAAATTGGAATTGAGAGAGCATGATGTCAGAGAAC	1922
Db	2044	AAGCAGTTTAAACGCGAAGAGCTGCCAAATTGGAATTGAGAGAGCATGATGTCAGAGAAC	2103
QY	1921	AGGTCCACGGGAAGAGCTGGGGGAAAGTGGGCACTGACTTATCTTTCCGGCAGCATGGA	1980
Db	2104	AGGTCCACGGGAAGAGCTGGGGGAAAGTGGGCACTGACTTATCTTTCCGGCAGCATGGA	2163
QY	1981	ATCATTTAGAGTCTCC	1995
Db	2164	ATCATTTAGAGTCTCC	2178

QY	61	AGTGGAA	CGGAAAAA	AGTGCTGCTTA	TGATAGCCGGCCATTTGTGGA	TACATACATCC	120
Db	244	AGTGGAA	CGGAAAAA	AGTGCTGCTTA	TGATAGCCGGCCATTTGTGGA	TACATACATCC	303
QY	121	CACATTTTGG	AGGCCATTATATCA	AGTCGTCCAGCTATATAG	GGGAAGTTGCAACG		180
Db	304	CACATTTTGG	AGGCCATTATATCA	AGTCGTCCAGCTATATAG	GGGAAGTTGCAACG		363
QY	181	GACAAAGTGT	TAAATTA	CAAGAGCTCA	CCAGCATTCAGCGAA	CATTAAGTTGACAT	240
Db	364	GACAAAGTGT	TAAATTA	CAAGAGCTCA	CCAGCATTCAGCGAA	CATTAAGTTGACAT	423
QY	241	TGCAGTCA	GAAAGGTGTGATTA	TGATCA	TCAAAGCTCCAA	GATGTGCTCTCTCTTCA	300
Db	424	TGCAGTCA	GAAAGGTGTGATTA	TGATCA	TCAAAGCTCCAA	GATGTGCTCTCTCTTCA	483
QY	301	GACGTGTTTC	ACACTGTA	CTTCGTGGGTAA	CTGGAGAAAGGTTCA	ACTCTGTCA	360
Db	484	GACGTGTTTC	ACACTGTA	CTTCGTGGGTAA	CTGGAGAAAGGTTCA	ACTCTGTCA	543
QY	361	CTTGGAGGTG	GGTTGTC	GAGTTCCTCTGTTGTTTCC	TGGGCTCTGTGGA	AGGAAATCC	420
Db	544	CTTGGAGGTG	GGTTGTC	GAGTTCCTCTGTTGTTTCC	TGGGCTCTGTGGA	AGGAAATCC	603
QY	421	ACTCTAGTCC	CTACCTGCAT	TTCTTCAGGCTTGCTTAC	CTGTGTGCCA	CAATTTGGGCCAAC	480
Db	604	ACTCTAGTCC	CTACCTGCAT	TTCTTCAGGCTTGCTTAC	CTGTGTGCCA	CAATTTGGGCCAAC	663
QY	481	CGAAATCTTCC	CAACTCTTATCT	TGGCTGGCGACG	CAAGATATGTCCTTAA	CAAGGATGTCAT	540

RESULT 11					
LOCUS	BD183422		4790 bp	DNA	linear
DEFINITION	Novel genes and proteins encoded by the genes.				
ACCESSION	BD183422				
VERSION	BD183422.1		GI:1875622		
KEYWORDS	JP 2002345492-A/135.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				

DB	Sequence	Position
DB	644 GGAATCTCTCCCAACTTATATCTGGCCGACGACGANGTCTCAACAAAGAGCGTGANG	723
QY	541 CAGCAGAAATGGAGTTGGTTATGTGTTAAATGCAAGCTATACCTGTGTCAAAGCGTCACTTT	600
DB	724 CAGCAGAAATGGAGTTGGTTATGTGTTAAATGCAAGCAATACTGTGCCAAAGCGTCACTTT	783
QY	601 ATCCCCGAGTCACATTTCTGCGGTGTGCTGTGAAATGACAGCTTTTGTGTGAGAAATTTTGG	660
DB	784 ATCCCCGAGTCACATTTCTGCGGTGTGCTGTGAAATGACAGCTTTTGTGTGAGAAATTTTGG	843

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 4790)	Ohara, O., Nagase, T. and Nakajima, D.	Novel genes and proteins encoded by the genes Patent: JP 2002345492-A 135 03-DEC-2002;	KAZUSA DNA RESEARCH INSTITUTE	OS Homo sapiens (human)

Db 844 CCGTGGTGGCAAAATCAGTAGATTCTTCATTGAGAAAGCAAAAGCCTCCAAATGATGTGTT 903

Qy 721 CTAGTGCACCTTTTAAGCTGGGATCTCCCGCTCCGCACACATCGCTATGCGCTCAATCATG 780

Db 904 CTAGTGCACCTTTTAAGCTGGGATCTCCCGCTCCGCACACATCGCTATGCGCTCAATCATG 963

FEATURES	source
PF	26-FEB-2002 JP 2002049009
PI	OSAMO OHARA, TAKAHIRO NAGASE, DAISUKE NAKAJIMA
PC	C12N15/09, C07K14/47//A61K31/711, A61K38/00, A61K48/00, A61P25/00,
PC	A61P25/14,
PC	A61P25/18, A61P35/00, C12N15/00, A61K37/02
CC	Novel genes and proteins encoded by the genes FH
Location/Qualifiers	
FT	CDS (184) . . (2178) .
	Location/Qualifiers
	1. .4790

Db	Seq	Accession
Db	664 AAGAGGATGGAACAATCTTTAGATGAAGCTTACAGATTTGTGAAAGAAAAAGACTCT	1023
Qy	841 ATTATCTCAAACTCAATTTTCTGGGCCCACTCTGGCATATGAGAAAGATTAAAGC	900
Db	1024 ATATCTCCAAACTCAATTTTCTGGGCCCACTCTGGCATATGAGAAAGATTAAAGC	1083
Qy	901 CAGACTGGAGCATCAGGGCCAAAGAGCAAACTCAAGCTGCGCACTTGAAGAGCAAT	960
Db	1084 CAGCTGAGAGCATCAGGGCCAAAGAGCAAACTCAAGCTGCGCACTTGAAGAGCAAT	1143

```

/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

```

Qy	961	GAACCTGTCCTGCTGTTCTCAGAGGGTGGACAGAAAAAGCGAGCCCTCAGTTCACCC	1020
Db	1144	GAACCTGTCCTGCTGTTCTCAGAGGGTGGACAGAAAAAGCGAGCCCTCAGTTCACCC	1203

Query Match	94.9%;	Score 1893;	DB 6;	Length 4790;
Best Local Similarity	99.9%;	Pred. No. 0;		
Matches 1993; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0

1021 TTGTCGCGACTCTCTACCTCAAGAGGCACACAGGCAAAAGCCCGCGCATCCGCGCAGCCGG 1080
1204 TGTGCGCGACTCTCTACCTCAAGAGGCACACAGGCAAAAGCCCGCGCATCCGCGCAGCCGG 1263

DY 1 ATGGCCCATGAGATGTGGAATAATTCTACCGAGGTTGGTGCCTCCTGAA 60

Db 184 ATGCCCATGAGATGTGGAATAATTCTACCGAGGTTGGTGCCTCCTGAA 243

Db 1264 CCAGCGCTGCCAGCGTGCAGCCGTGCTGTAGAGGACAGCCCGCTGGTACAGGCGCTC 1323

QY 1141 AGGAGGCTGACCTGTCCTCCGAGACAGGCTGGAAGACGATTAAGCTCAAGCGTTCCTTC 1200
 DB 1324 AGTGGGCTGCACCTGTCCTCCGAGACAGGCTGGAAGACGATTAAGCTCAAGCGTTCCTTC 1383
 QY 1201 TCTCTGGATATCAATCAATGATTTCAATATTCAGCCAGAGATGAGCATCTTACATGGCTTC 1260
 DB 1384 TCTCTGGATATCAATCAATGATTTCAATATTCAGCCAGAGATGAGCATCTTACATGGCTTC 1443
 QY 1261 TCTCTCATCAGAAATGCTTTGGAAATCTACAAACCTTCCACTCTCTGGATGGGACCAAC 1320
 DB 1444 TCTCTCATCAGAAATGCTTTGGAAATCTACAAACCTTCCACTCTCTGGATGGGACCAAC 1503
 QY 1321 AAGCTATGCGAGATTTCTCCCTGTTACAGAACTATGAGGAGAGACTCCCGAAACAGTCTCT 1380
 DB 1504 AAGCTATGCGAGATTTCTCCCTGTTACAGAACTATGAGGAGAGACTCCCGAAACAGTCTCT 1563
 QY 1381 GATTAAGAGAGAGACGACATCTCCCAAGAACTGACAGCCCGAGGCTTTCAGACAGCCAG 1440
 DB 1564 GATTAAGAGAGAGACGACATCTCCCAAGAACTGACAGCCCGAGGCTTTCAGACAGCCAG 1623
 QY 1441 AGCAAGGATTCATTCCTGGTCAAGACGAGACGAGTGGACCCCGACAGAGTCTCTTTTA 1500
 DB 1624 AGCAAGGATTCATTCCTGGTCAAGACGAGACGAGTGGACCCCGACAGAGTCTCTTTTA 1683
 QY 1501 TCTCCACTGCATGGAAGTGGGAGCGTGGAGAGCAATTACACACGAGCTTCTTTCGCGC 1560
 DB 1684 TCTCCACTGCATGGAAGTGGGAGCGTGGAGAGCAATTACACACGAGCTTCTTTCGCGC 1743
 QY 1561 CTTTCCACGACGACGACGACACTCAGCAAGTCTGCTGCGCTTGAAGGCTTGGAC 1620
 DB 1744 CTTTCCACGACGACGACGACACTCAGCAAGTCTGCTGCGCTTGAAGGCTTGGAC 1803
 QY 1621 TCGGATATCTTGGCCCCCAGACCTCTAACCCTTCCCTGACGACGCTGATTTTGGC 1680
 DB 1804 TCGGATATCTTGGCCCCCAGACCTCTAACCCTTCCCTGACGACGCTGATTTTGGC 1863
 QY 1681 ACAGAGCTCTCACAATCTTCTGCTGCTGACGATCTAGGAGGAGGAGTGCAGTACTCTC 1740
 DB 1864 ACAGAGCTCTCACAATCTTCTGCTGCTGACGATCTAGGAGGAGGAGTGCAGTACTCTC 1923
 QY 1741 GCCTACAGCTGACGACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1800
 DB 1924 GCCTACAGCTGACGACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1983
 QY 1801 CAGAAAGCAAGTGAAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1860
 DB 1984 CAGAAAGCAAGTGAAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2043
 QY 1861 AAGCAGTTTAAAGCGAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1920
 DB 2044 AAGCAGTTTAAAGCGAGAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2103
 QY 1921 AGGTCACGGAAGAGTGGGGAAGAGTGGGAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTG 1980
 DB 2104 AGGTCACGGAAGAGTGGGGAAGAGTGGGAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTG 2163
 QY 1981 ATCATTGAGGTCTCC 1995
 DB 2164 ATCATTGAGGTCTCC 2178
 RESULT 12
 AB051487 4790 bp mRNA linear PRI 07-FEB-2001
 LOCUS AB051487
 DEFINITION Homo sapiens mRNA for KIAA1700 protein, partial cds.
 ACCESSION AB051487
 VERSION AB051487.1 GI:12697944
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases)

AUTHORS Nagase, T., Kikuno, R., Hattori, A., Kondo, Y., Okumura, K. and Ohara, O.
 TITLE Prediction of the coding sequences of unidentified human genes. XIX. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro
 JOURNAL DNA Res. 7 (6), 347-355 (2000)
 MEDLINE 21082932
 PUBMED 11214970
 REFERENCE 2 (bases 1 to 4790)
 AUTHORS Ohara, O., Nagase, T. and Kikuno, R.
 TITLE Direct Submission
 JOURNAL Submitted (22-NOV-2000) Osamu Ohara, Kazusa DNA Research Institute, Department of Human Gene Research, 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: cdainfo@kazusa.or.jp, URL: http://www.kazusa.or.jp/huge, Tel: 81-438-52-3913, Fax: 81-438-52-3914)
 FEATURES
 source
 1..4790
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="FJ15353"
 /note="vector: pBluescriptII SK plus"
 1..4790
 /gene="KIAA1700"
 <109..2181
 /gene="KIAA1700"
 /note="Start codon is not identified."
 /codon_start=1
 /product="KIAA1700 protein"
 /protein_id="BAB21791.1"
 /db_xref="GI:12697945"
 /translation="APEHTTITSSQIQEGGKRGILVYMAHEMTCITVTRIVALL
 ESEIKVLLIDSRPFVETINTSHILEININCKMRRLQDQKVLITELIQSARKV
 DIDCSQKVVVYDQSSQDVASLSDCFLLVLKLEKSFVSLHLAGFAEFSRCPGL
 CEKGLVETICISQCLPVANIGPRLIPNLVLCQDVLNKLMOQNGIYVLANSN
 TCRPDLFIESHFRLPVPVDSFCEKILPMLDKSVDFIEKAKASNGCVLHCLAGISRS
 ATIAIAYIKRMDSLDEAYRFEVKEKRPITISNPNFNLGOLLVEKKIKNOTGAGSRS
 KIKLILAEKPNPVPVAVSGKQSRPLSPKADATSAAGORPHNPASVSPVSL
 PELEDSPLVQLSGILHLSADRLBESNKKRPSLDIKSVISASMAHLHGFSSSD
 ALEYYKSTTLDGINKLCQPSVQELSEQTPETSPDKESASIPKLIQTARPSDSQR
 LHSVTSSEGTQORSLLSPHRSQVEDNVTSLFGLSTSOQHLTKSAGLIGKWS
 DILAPQTSPLTSSWYFATESHFSASAIYGSASVAYSQSCSLPTCGDQVSVR
 ROKPDRADRSRSHWESPFKQFRRSCQMFESIMENSRBELGKVSQSSFSGS
 SNEILIEVS"
 ORIGIN
 Query Match 94.9%; Score 1893; DB 9; Length 4790;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 ATGGCCCATGAGTGAATGGAATCTCAATTTGTTACTGAGAGTGGTGGCTCTGCTGAA 60
 DB 184 ATGGCCCATGAGTGAATGGAATCTCAATTTGTTACTGAGAGTGGTGGCTCTGCTGAA 243
 QY 61 AGTGAACGGAAGAAAGTGTGCTAATGTAATGTAACCGGCGCATTTGTGAATACATATACATCC 120
 DB 244 AGTGAACGGAAGAAAGTGTGCTAATGTAATGTAACCGGCGCATTTGTGAATACATATACATCC 303
 QY 121 CACATTTTGAAGCCATATATATCAACTGCTCCAACTTATGAAAGGAGTTGCAACAG 180
 DB 304 CACATTTTGAAGCCATATATATCAACTGCTCCAACTTATGAAAGGAGTTGCAACAG 363
 QY 241 TGCAGTCAGAGTGTATGATTCATCAATCAAGCTCCCAAGATGGTCCCTCTCTTCA 300
 DB 424 TGCAGTCAGAGTGTATGATTCATCAATCAAGCTCCCAAGATGGTCCCTCTCTTCA 483
 QY 301 GACTGTTTCTCACTGTACTTCTGGGTAAGTGAAGAGGTTCAACTCTGTTCACTG 360
 DB 484 GACTGTTTCTCACTGTACTTCTGGGTAAGTGAAGAGGTTCAACTCTGTTCACTG 543

QY 361 CTTGAGGTGGGTTTCTGAGTTCTCTGTTTCCCTGGCTCTTGAAAGAAATCC 420
DB 544 CTTGAGGTGGGTTTCTGAGTTCTCTGTTTCCCTGGCTCTTGAAAGAAATCC 603
QY 421 ACTTAGTCCCTACCTGATTTCTGAGCTTGTCTTACCTGTTGCCAATGAGGCAAC 480
DB 604 ACTTAGTCCCTACCTGATTTCTGAGCTTGTCTTACCTGTTGCCAATGAGGCAAC 663
QY 481 GCAATTTCTCCCAATCTTTATTTGGCTGCGAGGAGATGTCCTCAACAGAGCTGTA 540
DB 664 GCAATTTCTCCCAATCTTTATTTGGCTGCGAGGAGATGTCCTCAACAGAGCTGTA 723
QY 541 CAGCAGAAATGGATGTTAATGTTAATGCGAGCTTACCTGTCACAAAGCTGACTT 600
DB 724 CAGCAGAAATGGATGTTAATGTTAATGCGAGCAATACCTGTCACAAAGCTGACTT 783
QY 601 ATCCCGAGTCTCATTTCTGCGGTGCTGTGAATGACAGCTTTTGTGAAAAATTTTG 660
DB 784 ATCCCGAGTCTCATTTCTGCGGTGCTGTGAATGACAGCTTTTGTGAAAAATTTTG 843
QY 661 CGGTGTTGGA CAATCAGTAGATTTCATGAGAAAGAAAGGCTTCATGATGTT 720
DB 844 CGGTGTTGGA CAATCAGTAGATTTCATGAGAAAGCAAAAGCTTCATGATGTT 903
QY 721 CTAAGTCACTGTTAGCTGGGATCTCCGCTCCGCAACATCGCTATCGCTACATG 780
DB 904 CTAAGTCACTGTTAGCTGGGATCTCCGCTCCGCAACATCGCTATCGCTACATG 963
QY 781 AAGAGGATGGA CAGTCTTTTATGAAAGCTTACAGATTTGTGAAAGAAAGCACTT 840
DB 964 AAGAGGATGGA CAGTCTTTTATGAAAGCTTACAGATTTGTGAAAGAAAGCACTT 1023
QY 841 ATATCTCAAACTTCAATTTTCTGGGCAACTCTGAGCTATGAGAAAGAAAGTTAAG 900
DB 1024 ATATCTCAAACTTCAATTTTCTGGGCAACTCTGAGCTATGAGAAAGAAAGTTAAG 1083
QY 901 CAGCTGAGGATGAGGCGCAAGAGCAAACTGAGCTGTCAGCTGAGAGCCAAAT 960
DB 1084 CAGCTGAGGATGAGGCGCAAGAGCAAACTGAGCTGTCAGCTGAGAGCCAAAT 1143
QY 961 GAACTGTCCCTGCTGCTGAGAGGTTGGA CAGAAAGGAGAGCGCCCTCACTGACCC 1020
DB 1144 GAACTGTCCCTGCTGCTGAGAGGTTGGA CAGAAAGGAGAGCGCCCTCACTGACCC 1203
QY 1021 TGTGCGACTGTGCTGCTGAGAGGAGAGCAAGAGCGCTGATCCGCGAGCGTG 1080
DB 1204 TGTGCGACTGTGCTGCTGAGAGGAGAGCAAGAGCGCTGATCCGCGAGCGTG 1263
QY 1081 CCGAGCTGCGGAGCTGAGCGCTGCTGTTAGAGGAGAGCGCGCTGATGAGCGCTG 1140
DB 1264 CCGAGCTGCGGAGCTGAGCGCTGCTGTTAGAGGAGAGCGCGCTGATGAGCGCTG 1323
QY 1141 AGTGGCTGCACTGTCCGAGAGAGGCTGGAAGAGAGAGAGAGAGGCTTCTTC 1200
DB 1324 AGTGGCTGCACTGTCCGAGAGAGGCTGGAAGAGAGAGAGAGAGGCTTCTTC 1383
QY 1201 TCTCTGATATCAATCAAGTTTCAATGAGCCAGAGTGCAGATCTTACATGCTTC 1260
DB 1384 TCTCTGATATCAATCAAGTTTCAATGAGCCAGAGTGCAGATCTTACATGCTTC 1443
QY 1261 TCTCTATCAAGAGATGCTTGAATATCAAACTTCACTGCTGATGAGGCAAC 1320
DB 1444 TCTCTATCAAGAGATGCTTGAATATCAAACTTCACTGCTGATGAGGCAAC 1503
QY 1321 AAGCTATGCAAGTTTCCCTGTTGAGAACTATGAGAGAGAGCTCCGGAACAGCTCT 1380
DB 1504 AAGCTATGCAAGTTTCCCTGTTGAGAACTATGAGAGAGAGCTCCGGAACAGCTCT 1563
QY 1381 GATTAAG 1440
DB 1564 GATTAAG 1623

QY 1441 AGCAGCGATGTCATTCGTTGAGAAACAGACAGTGGACCGCCAGAGTCCCTTTTA 1500
DB 1624 AGCAGCGATGTCATTCGTTGAGAAACAGACAGTGGACCGCCAGAGTCCCTTTTA 1683
QY 1501 TCTCCACTGATTCGAATGGAGAGCTGAGAGCAATTAACACACAGCTTCTTTGAG 1560
DB 1684 TCTCCACTGATTCGAATGGAGAGAGCTGAGAGCAATTAACACAGCTTCTTTGAG 1743
QY 1561 CTTTCACGAGCAGACAGACCTCA GAAAGTGTGCTGGCTGAGGCTTAAGGAGCTGAC 1620
DB 1744 CTTTCACGAGCAGACAGACCTCA GAAAGTGTGCTGGCTGAGGCTTAAGGAGCTGAC 1803
QY 1621 TGGATATTTGGCCCCCAGACTCTA CCGCTTCCCTGACAGCAGCTGATTTTGC 1680
DB 1804 TGGATATTTGGCCCCCAGACTCTA CCGCTTCCCTGACAGCAGCTGATTTTGC 1863
QY 1681 ACAGAGTCTCA CACTTCTA CTTGSCCTCAGCACTTAACGAGGACAGTCACTTCT 1740
DB 1864 ACAGAGTCTCA CACTTCTA CTTGSCCTCAGCACTTAACGAGGACAGTCACTTCT 1923
QY 1741 GCTTACAGTGCAGCAGCTGCCACTT GCGAGAGCAAGTCTATTCTGTGCGAGCG 1800
DB 1924 GCTTACAGTGCAGCAGCTGCCACTT GCGAGAGCAAGTCTATTCTGTGCGAGCG 1983
QY 1801 CAGAGCCAGTGCAGAGCTGACTT CCGGCGGAGAGCTGCGATTGAAGAGAGCCCTTTGAA 1860
DB 1984 CAGAGCCAGTGCAGAGCTGACTT CCGGCGGAGAGAGCTGCGATTGAAGAGAGCCCTTTGAA 2043
QY 1861 AAGCAGTTTAAACAGAGAGCTGCAATGGAATTTGAGAGACATCATGTGAGAAAC 1920
DB 2044 AAGCAGTTTAAACAGAGAGCTGCAATGGAATTTGAGAGACATCATGTGAGAAAC 2103
QY 1921 AGTCA CCGGAAAGCTGGGGAAGTGGGAGTCACTTACCTTTTGGGCAAGATGAA 1980
DB 2104 AGTCA CCGGAAAGCTGGGGAAGTGGGAGTCACTTACCTTTTGGGCAAGATGAA 2163
QY 1981 ATCATTGAGTCTCC 1995
DB 2164 ATCATTGAGTCTCC 2178

RESULT 13
AX180875 2732 bp DNA linear PART 06-AUG-2001
LOCUS AX180875
DEFINITION Sequence 2 from Patent WO0146394.
ACCESSION AX180875
VERSION AX180875.1 GI:15132703
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1. Ploeman, G.D., Martinez, R., Whyte, D., Manning, G., Sudarsanam, S.,
Hill, R.U., and Flanagan, P.
TITLE Mammalian protein phosphatases
JOURNAL Patent: WO 0146394-A 2 28-JUN-2001;
Sugen, Inc. (US)
FEATURES
Location/Qualifiers
source 1..2732
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Query Match 92.3%; Score 1842; DB 6; Length 2732;
Best local Similarity 99.8%; Pred. No. 0;
Matches 1992; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGGCCCATGATGATGTTGGAATCAATTTGTTACTGAGAGGTTGTGCTGTGAA 60
DB 538 ATGGCCCATGATGATGTTGGAATCAATTTGTTACTGAGAGGTTGTGCTGTGAA 597

TITLE Xue, A.-J., Yang, Y., Wehrman, T. and Drimanac, R.T.
Novel nucleic acids and polypeptides
Patent: WO 0222660-A 115 21-MAR-2002;
JOURNAL HUSBQ, INC. (US)

FEATURES
source Location/Qualifiers

CDS
1..3104
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
159..2156
/note="unamed protein product"
/codon_start=1
/protein_id="CAD34807.1"
/db_xref="GI:21438840"
/translation="MAHMLCTOIYTRIVALLBSGTEKULIDSRPFVYNTSHILE
AINTNSKLMKRRLOQDKVLTTELIOGSAKVIDDSQKVVYDSQVVASISPC
FLTVLGLKESFNSVHLAAGFAEFRCPEGLCEGSLTVPICISQPCLPVANIPT
RILPNLYGQDRVANKELMOQNGIGVLAISRSATIPESHFLRPVNDSPCEK
ILPWLDKSVDFIERAKASNGCVLHCLAGISRSATIPESHFLRPVNDSPCEK
KRPTISPFNFILQGLDYERKKIKNOTGASGPKSLKILHEKEMRPVAPRQKSE
TLPSPCADSATSEAGORPVHPASVPSVSVOPSLIEDSPYQALSGHLSADRLSD
SNKLRPSLIDKSVYSASMAASHFSSSEDALEYKPSSTLIDGNKLCQSPVOE
LSQTPETSPDKEXASTPKUQTARPSDSQKRLHSVRTSSGTAQRSLSPHRSQS
VEDNYHTSPLEGSTSQQLTKSAGLKGHMSDILAPQTSSTLSISWTFATESHP
YASAIYGGASVAYSCSQLPTCGDQVSVRRORRQADSRDRSWMHSESPERQFK
RRSCMEFSGSIMSENSREBLKVGSGSFSGSMELIEVS"

ORIGIN

Query Match 92.3%; Score 1842; DB 6; Length 3104;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1992; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGCCCATGATGATGATGAACTCAATGTTACTGAGAGGTGGTGGCTGCTGGGA 60
Db 159 ATGCCCATGATGATGATGAACTCAATGTTACTGAGAGGTGGTGGCTGCTGGGA 218
QY 61 AGTGGAAACGAAAAAGTGGCTGATTAATGATAGCCGGCCATTTGTGAAATACATACATCC 120
Db 219 AGTGGAAACGAAAAAGTGGCTGATTAATGATAGCCGGCCATTTGTGAAATACATACATCC 278
QY 121 CACATTTGGAGGCAATTAATACACTGCTCCAGCTTATGAGGAAAGTTGCAACG 180
Db 279 CACATTTGGAGGCAATTAATACACTGCTCCAGCTTATGAGGAAAGTTGCAACG 338
QY 181 GACAAAGTGTAAATTCAGAGCTCATCCAGCATTCAGCGAAACATTAAGTTGACATGAT 240
Db 339 GACAAAGTGTAAATTCAGAGCTCATCCAGCATTCAGCGAAACATTAAGTTGACATGAT 398
QY 241 TGCAGTCAGAAAGTTGATTAAGATCAAAAGCTCCCAAGATGTTGCTCTCTCTTCA 300
Db 399 TGCAGTCAGAAAGTTGATTAAGATCAAAAGCTCCCAAGATGTTGCTCTCTCTTCA 458
QY 301 GACTGTTTTCTCACTGATCTTGGGTAACCTGSAAGAGAGCTTCAACTCTGTTCACTG 360
Db 459 GACTGTTTTCTCACTGATCTTGGGTAACCTGSAAGAGAGCTTCAACTCTGTTCACTG 518
QY 361 CTTCGAGGTGGGTTGCTGAGTCTCTCGTTGTTTCCCTGGCTCTGTGAAGAAATCC 420
Db 519 CTTCGAGGTGGGTTGCTGAGTCTCTCGTTGTTTCCCTGGCTCTGTGTGAAGAAATCC 578
QY 421 ACTCTAGTCCCTACCTGCACTTCTCAGCTTGGCTTACCTGTTGCCAATTTGGGCAAC 480
Db 579 ACTCTAGTCCCTACCTGCACTTCTCAGCTTGGCTTACCTGTTGCCAATTTGGGCAAC 638
QY 481 CGAATTTCTCCCAATCTTATCTTGGCTGCGAGGAGATGCTCAAGAGAGCTGATA 540
Db 639 CGAATTTCTCCCAATCTTATCTTGGCTGCGAGGAGATGCTCAAGAGAGCTGATA 698
QY 541 CAGCGAATGGGATGTTGTTAATGCGACCTATACCTGTCAGAAAGCTGACTTT 600
Db 699 CAGCGAATGGGATGTTGTTAATGCGACCTATACCTGTCAGAAAGCTGACTTT 758

QY 601 ATCCCGAGTCTCATTTCTCGGTGCTGCTGAAATGACAGCTTTGTGAGAAAATTTTG 660
Db 759 ATCCCGAGTCTCATTTCTCGGTGCTGCTGAAATGACAGCTTTGTGAGAAAATTTTG 818
QY 661 CCGTGGTTGGACAAATCACTAATTTCAATTGAGAAAAGCAAAAGCTCCAAATGATGTTG 720
Db 819 CCGTGGTTGGACAAATCACTAATTTCAATTGAGAAAAGCAAAAGCTCCAAATGATGTTG 878
QY 721 CTAGTGACATGTTTAGTGGATCTCCGCTCCGCAACATGCTATGCTCATCATG 780
Db 879 CTAGTGACATGTTTAGTGGATCTCCGCTCCGCAACATGCTATGCTCATCATG 938
QY 781 AAGAGATGACATGCTTTAATGAGTTACAGATTTGTGAAAAGAAAGACTTACT 840
Db 939 AAGAGATGACATGCTTTAATGAGTTAAGATTTGTGAAAAGAAAGACTTACT 998
QY 841 ATATCTCCAACTTCAATTTTCTGGGCAACCTCCGCACTATGAGAAAGATTAAGAC 900
Db 999 ATATCTCCCACTTCAATTTTCTGGGCAACCTCCGCACTATGAGAAAGATTAAGAC 1058
QY 901 CAGCTGAGCATCAGGCGCAAAAGCAAACTCAAGCTGTCACCTGAGAAAGCAAAAT 960
Db 1059 CAGCTGAGCATCAGGCGCAAAAGCAAACTCAAGCTGTCACCTGAGAAAGCAAAAT 1118
QY 961 GAACTGTCCCTGCTGTCTCAGAGGTGGAAGAAAAGCGAGAGCCCTCAGTCCACC 1020
Db 1119 GAACTGTCCCTGCTGTCTCAGAGGTGGAAGAAAAGCGAGAGCCCTCAGTCCACC 1178
QY 1021 TGTGCGCATCTGCTACCTCAGAGGCAAGCAAAAGCCCGTGCAATCCCGCAAGGTG 1080
Db 1179 TGTGCGCATCTGCTACCTCAGAGGCAAGCAAAAGCCCGTGCAATCCCGCAAGGTG 1238
QY 1081 CCCAGCTGCGCAGGTCGACCCCTGCTGTTAAGAGACAGCCGCTGATACAGCGCTC 1140
Db 1239 CCCAGCTGCGCAGGTCGACCCCTGCTGTTAAGAGACAGCCGCTGATACAGCGCTC 1298
QY 1141 AGTGGGCTGCACTGTCCGCAAGAGCTGGAACACAGCAATAGCTTCAAGCTTCTTC 1200
Db 1299 AGTGGGCTGCACTGTCCGCAAGAGCTGGAACACAGCAATAGCTTCAAGCTTCTTC 1358
QY 1201 TCTCTGATTAATCAATCAGTTTCAATTAATTCAGCCAGATGCGATCTTACATGCTTC 1260
Db 1359 TCTCTGATTAATCAATCAGTTTCAATTAATTCAGCCAGATGCGATCTTACATGCTTC 1418
QY 1261 TCTCTATCAAGAAATGCTTTGGAATTAATCAAACTTCACTACTGTGATGAGCAAC 1320
Db 1419 TCTCTATCAAGAAATGCTTTGGAATTAATCAAACTTCACTACTGTGATGAGCAAC 1478
QY 1321 AAGCTATGCAAGTTCTCCCTGTTCAAGAACTATGGAAGAGACTCCGAAACCAAGTCT 1380
Db 1479 AAGCTATGCAAGTTCTCCCTGTTCAAGAACTATGGAAGAGACTCCGAAACCAAGTCT 1538
QY 1381 GATTAAGAGAAAGCAGCATCTCCCAAGAACTGACAGCGCCAGGCTTACAGACACAG 1440
Db 1539 GATTAAGAGAAAGCAGCATCTCCCAAGAACTGACAGCGCCAGGCTTACAGACACAG 1598
QY 1441 AGCAAGGATGATGATTCGCTCAAGAACCAAGAGATGCGACCGCCAGAGTCCCTTTA 1500
Db 1599 AGCAAGGATGATGATTCGCTCAAGAACCAAGAGATGCGACCGCCAGAGTCCCTTTA 1658
QY 1501 TCTCACTGATGAGATGGAAGTGAAGCAATTAACCAACCAAGCTTCTCTTTGGG 1560
Db 1659 TCTCACTGATGAGATGGAAGTGAAGCAATTAACCAACCAAGCTTCTCTTTGGG 1718
QY 1561 CTTTCCACAGCAGCAGACCTTCAAGAACTGCTGAGCTGAGCTTAAAGGCTGCGAC 1620
Db 1719 CTTTCCACAGCAGCAGACCTTCAAGAACTGCTGAGCTGAGCTTAAAGGCTGCGAC 1778
QY 1621 TCGGATATCTGGGCCCCCAAGCTCTACCTTCCCTGACAGAGCTGATATTTGGC 1680
Db 1779 TCGGATATCTGGGCCCCCAAGCTCTTACCTTCCCTGACAGAGCTGATATTTGGC 1838
QY 1681 ACAGAGTCTCAACTTACTGCTCAGGCACTACAGAGGAGTGCAGTTACTCT 1740

```

Db      1839  AGAGATCTCCACACTTCTACTGCTCAGCCATCTAGGAGGAGCTGCAAGTTACTCT 1898
Qy      1741  GCTTACAGCTGCAGCCAGCTGCCCACTTGGGAGACCAGTCTATTCTGTGGCAGCG 1800
Db      1899  GCTTACAGCTGCAGCCAGCTGCCCACTTGGGAGACCAGTCTATTCTGTGGCAGCG 1958
Qy      1801  CAGAACCCAGTACAGAGCTGCTGCGCGGAGCTGGCATGAAGAGCCCTTTGA 1860
Db      1959  CAGAACCCAGTACAGAGCTGCTGCGCGGAGCTGGCATGAAGAGCCCTTTGA 2018
Qy      1861  AAGCAGTTTAAAGCAGAGCTGCCAAATGGAATTTGAGAGAGCATCATGTGAGAG 1920
Db      2019  AAGCAGTTTAAAGCAGAGAGCTGCCAAATGGAATTTGAGAGAGCATCATGTGAGAG 2078
Qy      1921  AGGTACCGGAGAGAGCTGGGAAAAGTGGGCGATGCTAGCTTTTTCGGGACAGTGA 1980
Db      2079  AGGTACCGGAGAGAGCTGGGAAAAGTGGGCGATGCTAGCTTTTTCGGGACAGTGA 2138
Qy      1981  ATGATTGAGCTCTCC 1995
Db      2139  ATGATTGAGCTCTCC 2153

RESULT 15
LOCUS      AX713989                2102 bp      DNA      linear      PAT 15-APR-2003
DEFINITION Sequence 673 from Patent EP1293569.
ACCESSION  AX713989
VERSION     AX713989.1  GI:29888917
KEYWORDS
SOURCE      Homo sapiens (human)
            Homo sapiens
            BUKARYOJA; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1
            Isogai,T., Sugiyama,T., Otsuki,T., Makamatsu,A., Sato,H., Ishii,S.,
            Yamamoto,J.I., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,
            Tamehika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahari,K. and
            Masuno,Y.
            Full-length cDNA
            Patent: EP 1293569-A 673 19-MAR-2003;
            Helix Research Institute (JP) ; Research Association for
            Biotechnology (JP)
FEATURES
            source
                1..2102
                /organism="Homo sapiens"
                /mol_type="unassigned DNA"
                /db_xref="taxon:9606"
ORIGIN
Query Match      89.8%; Score 1791; DB 6; Length 2102;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1991; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy      1  ATGGCCCATGATGATGATGGAATCAATGTTACTGAGAGGTTGGTGTCTGTGGAA 60
Db      56  ATGGCCCATGATGATGATGGAATCAATGTTACTGAGAGGTTGGTGTCTGTGGAA 115
Qy      61  AGTGAACGGAAGAAAGTCTGTAATTGATAGCCGCGCATTTGGAATCAATATACCC 120
Db      116  AGTGAACGGAAGAAAGTCTGTAATTGATAGCCGCGCATTTGGAATCAATATACCC 175
Qy      121  CACATTTTGGAGCCATTAATATCACTGCTCCAGGTTATGAAGGAAAGTTGCAACAG 180
Db      176  CACATTTTGGAGCCATTAATATCACTGCTCCAGGTTATGAAGGAAAGTTGCAACAG 235
Qy      181  GACAAAGTGTATATACAGAGCTCATCAGCATTCAGCGAAACATAAGTTGACATTGAT 240
Db      236  GACAAAGTGTATATACAGAGCTCATCAGCATTCAGCGAAACATAAGTTGACATTGAT 295
Qy      241  TGCAGTCAGAAAGTGTATATAGATCAAAAGCTCCCAAGATGTGCTCTCTCTCA 300

```

```

Db      296  TGCAGTCAGAAAGTGTATATAGATCAAAAGCTCCAAAGATGTTGCTCTCTCTTCA 355
Qy      301  GACTGTTTCTACATGTAATTTCTGGGTAAACTGGAAGAGAGTTAACTCTTACCTG 360
Db      356  GACTGTTTCTACATGTAATTTCTGGGTAAACTGGAAGAGAGTTAACTCTTACCTG 415
Qy      361  CTTGCAAGTGGGTTTGTAGTCTCTGTTTTCCTGGGCTCTGTGAAGAAATTC 420
Db      416  CTTGCAAGTGGGTTTGTAGTCTCTGTTTTCCTGGGCTCTGTGAAGAAATTC 475
Qy      421  ACTTATGTCCTTACCTGATTTCTCAGCCTTGCTTACTGTTGCCAAATTTGGCCACC 480
Db      476  ACTTATGTCCTTACCTGATTTCTCAGCCTTGCTTACTGTTGCCAAATTTGGCCACC 535
Qy      481  CGAATTTCTCCAAATCTTATATCTTGGCTGCCAGAGATGTTCTCAAGAGAGCTGAT 540
Db      536  CTAAATTCCTCCAAATCTTATATCTTGGCTGCCAGAGATGTTCTCAAGAGAGCTGAT 595
Qy      541  CAGCAAGATGGGATTTGTTATGTTAAATGACAGCTATACCTGTCCAAAGCCTGACTT 600
Db      596  CAGCAAGATGGGATTTGTTATGTTAAATGACAGCTATACCTGTCCAAAGCCTGACTT 655
Qy      601  ATCCCGAGTCTCATTTCTGCGGTGCTGTGAATGACAGCTTTTGTGAGAAATTTTG 660
Db      656  ATCCCGAGTCTCATTTCTGCGGTGCTGTGAATGACAGCTTTTGTGAGAAATTTTG 715
Qy      661  CCGTGGTTGCAAAATCAGTATGATTTTATTTGAAAGCAAAAGCCTCCAAATGATGTT 720
Db      716  CCGTGGTTGCAAAATCAGTATGATTTTATTTGAAAGCAAAAGCCTCCAAATGATGTT 775
Qy      721  CTATGCACTGTTTATGCTGGGATCTCCGCTCCGACCATGCTATGCTTACATCATG 780
Db      776  CTATGCACTGTTTATGCTGGGATCTCCGCTCCGACCATGCTTATGCTTACATCATG 835
Qy      781  AAGAGATGAGATGCTTATAGATGAGCTTACAGATTTGTGAAGAAAGAAAGCACTACT 840
Db      836  AAGAGATGAGATGCTTATAGATGAGCTTACAGATTTGTGAAGAAAGAAAGCACTACT 895
Qy      841  ATATCTCCAAATCTCAATTTTCTGGGCAACTCTGTGATATGAGAAAGATTAAGAAC 900
Db      896  ATATCTCCAAATCTCAATTTTCTGGGCAACTCTGTGATATGAGAAAGATTAAGAAC 955
Qy      901  CAGACTGAGCATCGGGCCAAAGCAAACTCAAGCTGCTGCACTGAGAGAGCCAAAT 960
Db      956  CAGACTGAGCATCGGGCCAAAGCAAACTCAAGCTGCTGCACTGAGAGAGCCAAAT 1015
Qy      961  GAACCTGTCTCTGTCTCTCAGAGGATGAGCAAGAAAGCGAGAGCCCTCACTGTCACC 1020
Db      1016  GAACCTGTCTCTGTCTCTCAGAGGATGAGCAAGAAAGCGAGAGCCCTCACTGTCACC 1075
Qy      1021  TGTGCCGACTGTGCTACTCAAGGAGCAGAGCAAAAGGCGGTGATCCCGCAGCGTG 1080
Db      1076  TGTGCCGACTGTGCTACTCAAGGAGCAGAGCAAAAGGCGGTGATCCCGCAGCGTG 1135
Qy      1081  CCCAGGTGCCAGGCTGAGCGCTGCTGTAGAGGACAGCCGCTGTACAGGGGCTC 1140
Db      1136  CCCAGGTGCCAGGCTGAGCGCTGCTGTAGAGGACAGCCGCTGTGTACAGGGGCTC 1195
Qy      1141  AGTGGGCTCACCTGTCCGACAGAGGCTGGAAGACAGCAATAGCTCAAGGCTTCTTC 1200
Db      1196  AGTGGGCTCACCTGTCCGACAGAGGCTGGAAGACAGCAATAGCTCAAGGCTTCTTC 1255
Qy      1201  TCTTGGATATCAATCAATTTTATATTCAGCCAGAGATGAGCATCTTACATGGCTTC 1260
Db      1256  TCTTGGATATCAATCAATTTTATATTCAGCCAGAGATGAGCATCTTACATGGCTTC 1315
Qy      1261  TCTTATCAGAGATGCTTTGGAATATCTCAAACTTCCATCACTGTGATGGGACCAAC 1320
Db      1316  TCTTATCAGAGATGCTTTGGAATATCTCAAACTTCCATCACTGTGATGGGACCAAC 1375
Qy      1321  AAGCTATGCAATTTCTCCCTGTTCAGAACTATCGAGACAGATCCGAAACCAATGCT 1380
Db      1376  AAGCTATGCAATTTCTCCCTGTTCAGAACTATCGAGACAGATCCGAAACCAATGCT 1435

```



```
QY 1381 GATTAAGGAGGAGCCAGCATCCCAAGAGCTGCAAGCCAGGCCCTTCAGACAGCCAG 1440
    |||||
Db 1436 GATTAAGGAGGAGCCAGCATCCCAAGAGAGCTGCAAGCTGCCAGGCCCTTCAGACAGCCAG 1495
    |||||
QY 1441 AGCAAGGATTTGATTTGGATCAGAACCCAGCAGCAGTGGCACCGCCCAAGAGTCCCTTTTA 1500
    |||||
Db 1496 AGCAAGGATTTGATTTGGATCAGAACCCAGCAGCAGTGGCACCGCCCAAGAGTCCCTTTTA 1555
    |||||
QY 1501 TCTCCACTGATGGAATGGGAGGCTGGAGAGCAATTAACACACAGCTTCCTTTTGGGC 1560
    |||||
Db 1556 TCTCCACTGATGGAATGGGAGGCTGGAGAGCAATTAACACACAGCTTCCTTTTGGGC 1615
    |||||
QY 1561 CTTTCCACAGCCAGCAGCAGCTCAAGAGTCTGCTGGCCCTGGGCCCTTAAGGGCTGGCAC 1620
    |||||
Db 1616 CTTTCCACAGCCAGCAGCAGCTCAAGAGTCTGCTGGCCCTTAAGGGCTGGCAC 1675
    |||||
QY 1621 TCGGATATCTTGGCCCCCAGACCTCTACCCCTTCCCTGACACAGCAGTGTATTTTGCC 1680
    |||||
Db 1676 TCGGATATCTTGGCCCCCAGACCTCTACCCCTTCCCTGACACAGCAGTGTATTTTGCC 1735
    |||||
QY 1681 ACAGAGTCTCAACATTTCTACTGCTCAGCCATCTACGAGGCAGTGCCAGTTACTCT 1740
    |||||
Db 1736 ACAGAGTCTCAACATTTCTACTGCTCAGCCATCTACGAGGCAGTGCCAGTTACTCT 1795
    |||||
QY 1741 GCCTAAGCTGCAGCCAGCTGCCCACTTGGGAGACCAAGTCTATTCTGTGGCAGGCGG 1800
    |||||
Db 1796 GCCTAAGCTGCAGCCAGCTGCCCACTTGGGAGACCAAGTCTATTCTGTGGCAGGCGG 1855
    |||||
QY 1801 CAGAGCCCAAGTACAGAGCTGACTGCGGCGGAGCTGGCATGAGAGAGCCCTTTGAA 1860
    |||||
Db 1856 CAGAGCCCAAGTACAGAGCTGACTGCGGCGGAGCTGGCATGAGAGAGCCCTTTGAA 1915
    |||||
QY 1861 AAGCAGTTTAAAGCAGAGCTGCCAATGGAATTTGAGAGAGCATGTCTCAGAGAAC 1920
    |||||
Db 1916 AAGCAGTTTAAAGCAGAGCTGCCAATGGAATTTGAGAGAGCATGTCTCAGAGAAC 1975
    |||||
QY 1921 AGGTCAAGGGAAGAGCTGGGGAAGTGGGAGTCAAGTCTAGCTTTTCGGGAGCATGAA 1980
    |||||
Db 1976 AGGTCAAGGGAAGAGCTGGGGAAGTGGGAGTCAAGTCTAGCTTTTCGGGAGCATGAA 2035
    |||||
QY 1981 ATCATTGAGGTCTCC 1995
    |||||
Db 2036 ATCATTGAGGTCTCC 2050
    |||||
```

Search completed: June 22, 2004, 07:13:19
Job time : 12419 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2004, 02:52:01 ; Search time 7674 Seconds

(without alignments)
7763.233 Million cell updates/sec

Title: US-10-029-345A-108_COPY_538_2532

Perfect score: 1995
Sequence: 1 atggcccatgagatgatg9.....tggaaatcatgaggtctcc 1995

Scoring table: OLIGO-NUC
Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0

Total number of hits satisfying chosen parameters: 200554

Minimum DB seq length: 1419

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estcom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vit:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_pig:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	53	2.7	3325	11	AK035652 Mus muscu
2	39	2.0	2573	11	AK014441 Mus muscu
3	19	1.0	1428	29	AY412468 Pan trogl
4	19	1.0	1449	29	AY412467 Homo sapi

C	5	19	1.0	1453	11	AK036059 Mus muscu
C	6	19	1.0	2013	29	AY418586 Mus muscu
C	7	19	1.0	2207	11	BC038231 Mus sapi
C	8	19	1.0	2289	29	AY410969 Mus muscu
C	9	19	1.0	2391	11	AK014482 Mus muscu
C	10	19	1.0	2985	11	AK040537 Mus muscu
C	11	18	0.9	1447	12	BM926476 Mus muscu
C	12	18	0.9	1506	12	BG118105 Mus muscu
C	13	18	0.9	1545	28	CC206026 Mus muscu
C	14	18	0.9	1564	12	BG757312 Mus muscu
C	15	18	0.9	2647	11	AK086250 Mus muscu
C	16	18	0.9	2742	11	AK047025 Mus muscu
C	17	18	0.9	3109	11	AK037848 Mus muscu
C	18	18	0.9	3185	11	AK048870 Mus muscu
C	19	18	0.9	3235	11	AK049360 Mus muscu
C	20	18	0.9	3386	11	AK035947 Mus muscu
C	21	18	0.9	3813	29	AY412149 Mus sapi
C	22	18	0.9	3813	29	AY412150 Pan trogl
C	23	18	0.9	4286	11	AK045549 Mus muscu
C	24	18	0.9	5769	29	AY412182 Mus muscu
C	25	18	0.9	5769	29	AY412183 Pan trogl
C	26	18	0.9	6013	11	BC046934 Mus sapi
C	27	17	0.9	1445	10	BE422076 HMM019CF
C	28	17	0.9	1450	28	CC236993 CH261-191
C	29	17	0.9	1454	11	BC000940 Homo sapi
C	30	17	0.9	1459	29	AY401655 Mus sapi
C	31	17	0.9	1471	12	B0054048 AGENDOUT
C	32	17	0.9	1496	28	CC237048 Mus sapi
C	33	17	0.9	1502	8	BC012770 Homo sapi
C	34	17	0.9	1504	29	AY411641 Mus muscu
C	35	17	0.9	1528	11	AK017629 Mus muscu
C	36	17	0.9	1580	11	AK033751 Mus muscu
C	37	17	0.9	1587	28	CC189276 CH261-17J
C	38	17	0.9	1589	29	AY403911 Homo sapi
C	39	17	0.9	1593	14	CF110912 Shultzom
C	40	17	0.9	1607	12	BC566227 AG58840 Pan trogl
C	41	17	0.9	1614	29	AG158840 Pan trogl
C	42	17	0.9	1618	11	AK075981 Mus muscu
C	43	17	0.9	1635	29	AY402679 Pan trogl
C	44	17	0.9	1640	12	B1552587 Mus sapi
C	45	17	0.9	1683	11	AK078482 Mus muscu

ALIGNMENTS

RESULT 1	AK035652	3325 bp	RNA	linear	HTC 19-SEP-2003
LOCUS	AK035652				
DEFINITION	Mus musculus adult male urinary bladder cDNA, RIKEN full-length enriched library, clone:9530081F05 product:TRUNCATED MRPK PHOSPHATASE 7 homolog [Homo sapiens], full insert sequence.				
ACCESSION	AK035652	GI:26330815			
VERSION	AK035652.1				
SOURCE	HTC; CAP trapper.				
ORGANISM	Mus musculus (house mouse)				
REFERENCE	1				
AUTHORS	Carinci, P. and Hayashizaki, Y.				
TITLE	High-efficiency full-length cDNA cloning				
JOURNAL	Methods. Enzymol. 303, 19-44 (1999)				
MEDLINE	99279253				
PUBMED	10349636				
REFERENCE	2				
AUTHORS	Carinci, P., Shibata, Y., Hayata, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.				
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes				
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)				
MEDLINE	20493374				
PUBMED	11042159				

REFERENCE
AUTHORS
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Komoto, H., Akiyama, J., Nishi, K., Kitsuina, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matakiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL
MEDLINE
PUBMED
20530913
11076861

TITLE
4
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

JOURNAL
REFERENCE
AUTHORS
5
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

JOURNAL
REFERENCE
AUTHORS
6
(bases 1 to 3325)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sugabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akanira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toyota, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Chemical Research, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-fes@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

TITLE
JOURNAL
COMMENT
7
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/
Location/Qualifiers
1. 3325
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:9530081F05"
/db_xref="MG1:2359973"
/db_xref="taxon:10090"
/clone="9530081F05"
/sex="male"
/tissue_type="urinary bladder"
/clone_1ib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
460..1080
/note="unnamed protein product; TRUNCATED MAPK PHOSPHATASE 7 homolog [Homo sapiens] (SPTK) [Q960S2, evidence: FASTV, 96.7%id, 86%length, match=369]
putative"
/codon_start=1
/protein_id="BAC29138.1"

ORIGIN
Query Match 2.7%; Score 53; DB 11; Length 3325;
Best Local Similarity 100.0%; Pred. No. 4, 2e-15;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 502 TTGGTGCGCTCTGCTGGAAGTGAACGAAAAAGTCTGCTTAATGATACGCC 95
43 TTGGTGCGCTCTGCTGGAAGTGAACGAAAAAGTCTGCTTAATGATACGCC 95
|||||
502 TTGGTGCGCTCTGCTGGAAGTGAACGAAAAAGTCTGCTTAATGATACGCC 554
|||||
/db_xref="GI:26330816"
/translation="MAHEMI GFOITVESIVALLSEGTETKVLIDSRPPVEYNTSHILE
AININCSKIMKRRLOODKVLITELIOHSAKHVDIDCNOKVYVNOSSDQCSLSDC
PLVILGKERSFPNSVHLLAGGAFRSCFPELCGSKSTLPVTQISQCLPVAIVIGPT
RLPPLNYLACQCDVANKRGASGHSLSCLSVASGKAAVHYIWSNLT"

RESULT 2
AK014441
LOCUS
DEFINITION
AK014441.1 GI:12852294
mus musculus 18 days pregnant adult female placenta and extra
embryonic tissue cDNA, RIKEN full-length enriched library,
clone:3830417M17 product:TRUNCATED MAPK PHOSPHATASE 7 homolog [Homo
sapiens], full insert sequence.
ACCESSION
VERSION
AK014441.1 GI:12852294
KEYWORDS
HTC; CAP trapper.
SOURCE
MUS MUSCULUS (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS
1
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
PUBMED
10349636

TITLE
JOURNAL
MEDLINE
PUBMED
20499374
11042159

REFERENCE
AUTHORS
2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komoto, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

TITLE
JOURNAL
MEDLINE
PUBMED
20530913
11076861

TITLE
JOURNAL
REFERENCE
AUTHORS
3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Komoto, H., Akiyama, J., Nishi, K., Kitsuina, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matakiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

TITLE
JOURNAL
MEDLINE
PUBMED
20530913
11076861

TITLE
JOURNAL
REFERENCE
AUTHORS
4
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

TITLE
JOURNAL
REFERENCE
AUTHORS
5
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

JOURNAL
REFERENCE
AUTHORS
6
(bases 1 to 2573)
Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hayatsu, N., Hiramoto, K., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hirooka, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,

```

FEATURES
Source
Location/Qualifiers
1..2573
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:3830417M17"
/db_xref="MGI:1903116"
/db_xref="Ensembl:ENSMUSG00000000000"
/clone="3830417M17"
/sex="female"
/tissue_type="placenta and extra embryonic tissue"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="18 days pregnant adult"
1..2573
/note="(RUNCABD MAPK PHOSPHATASE 7 homolog (Homo sapiens)) (SPR1Q26QS2, evidence: FASTV, 96.7%ID, 86%length, match=369)"

```

REFERENCE	1 (bases 1 to 1428)
AUTHORS	Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A., Todd,M.A., Tanenbaum,D.M., Civeello,D.R., Lu,F., Murphy,B., Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
TITLE	Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios
JOURNAL	Science 302 (5652), 1960-1963 (2003)
PUBMED	14671302
REFERENCE	2 (bases 1 to 1428)
AUTHORS	Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A., Todd,M.A., Tanenbaum,D.M., Civeello,D.R., Lu,F., Murphy,B., Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
TITLE	Direct Submission Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA
COMMENT	This sequence was made by sequencing genomic exons and ordering them based on alignment.
FEATURES	Location/Qualifiers 1..1428 /organism="Pan troglodytes" /mol_type="genomic DNA" /db_xref="taxon:9598" <1..>1428 /gene="DUSP10" /locus_tag="HCM4567"
ORIGIN	
Query Match	1.0%; Score 19; DB 29; Length 1428;
Best Local Similarity	100.0%; Pred. No. 1.1e+03;
Matches	19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Dn	745 TCCCGCTCCGCACCATCG 763 1237 TCCTGCTCCGCCACCATCG 1255
RESULT 4	
LOCUS	AY12467 1449 bp DNA linear GSS 16-DEC-2003
DEFINITION	Homo sapiens DUSP10 gene, VIRUAL TRANSCRIPT, partial sequence,
ACCESSION	AY12467
VERSION	AY12467.1 GI:39768432
KEYWORDS	GSS.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 1449) Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A., Todd,M.A., Tanenbaum,D.M., Civeello,D.R., Lu,F., Murphy,B., Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
REFERENCE	2 (bases 1 to 1449)
AUTHORS	Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarival,A., Todd,M.A., Tanenbaum,D.M., Civeello,D.R., Lu,F., Murphy,B., Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Sninsky,J.J., Adams,M.D. and Cargill,M.
TITLE	Direct Submission Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA
COMMENT	This sequence was made by sequencing genomic exons and ordering them based on alignment.
FEATURES	Location/Qualifiers 1..1449 /organism="Homo sapiens" /mol_type="genomic DNA"
source	

```

gene
    /db_xref="taxon:9606"
    <1..>1449
    /gene="DUSP10"
    /locus_tag="HCM4567"

ORIGIN
Query Match
Best Local Similarity 100.0%; Pred. No. 1.1e+03; Length 1449;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy
745 TCCCGTCGCCACCATCG 763
1237 TCCCGTCGCCACCATCG 1255

RESULT 5
AK036059/c
LOCUS
DEFINITION
AK036059 1453 bp mRNA linear HTC 19-SEP-2003
Mus musculus 16 days neonate cerebellum cDNA, RIKEN full-length
enriched library, clone:9630031G17
product:6-phosphogluconolactonase, full insert sequence.
ACCESSION
AK036059.1 GI:26331123
VERSION
HTC, CAP trapper.
KEYWORDS
Mus musculus (house mouse)
SOURCE
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
1 Carninci, P. and Hayashizaki, Y.
TITLE
High-efficiency full-length cDNA cloning
JOURNAL
Meth. Enzymol. 303, 19-44 (1999)
MEDLINE
99279253
PUBMED
10349636

REFERENCE
AUTHORS
2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL
Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE
20499374
PUBMED
11042159

REFERENCE
AUTHORS
3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Kono, H., Akiyama, J., Nishi, K., Kitsuini, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujimake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matshiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
TITLE
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL
Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE
20530913
PUBMED
11076861

REFERENCE
AUTHORS
4 The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
TITLE
Functional annotation of a full-length mouse cDNA collection
JOURNAL
Nature 409, 685-690 (2001)
MEDLINE
11076861
PUBMED
11076861

REFERENCE
AUTHORS
5 The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
TITLE
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL
Nature 420, 563-573 (2002)
MEDLINE
11076861
PUBMED
11076861

JOURNAL
REFERENCE
AUTHORS
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shitaki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
TITLE
Direct Submision
JOURNAL
Submitted (16-JUL-2001) Yoshinide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
COMMENT
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.
FEATURES
source
1..1453
location/Qualifiers
organism="Mus musculus"
mol_type="mRNA"
strain="C57BL/6J"
/db_xref="PANTOM DB:9630031G17"
/db_xref="MGI:2400432"
/db_xref="taxon:10090"
/clone="9630031G17"
/tissue_type="cerebellum"
/clone_id="RIKEN full-length enriched mouse cDNA library"
/dev_stage="16 days neonate"
35..682
/note="unnamed protein product: 6-phosphogluconolactonase
(MGI:MGI:1913421, GB|NM_025336, evidence: BLASTN, 100%,
match=352)
putative"
/codon_start=1
/protein_id="BAC29292.1"
/db_xref="GI:26331124"
/translation="MAAPPSLSVSSRQELGASLQVQAPASCSLEGDRGFRALG
LSPDSVSLNARLPPLAAPAPGAPARATLGCDFRVPFPAESTVGLVYTHLISK
LPSPDSVSLNARLPPLAAPAPGAPARATLGCDFRVPFPAESTVGLVYTHLISK
DHPLEDEREKTIVAPISDSRPPQRTVLTLPVINAQSIIIPVATSGKAAVAKVR"

ORIGIN
Query Match
Best Local Similarity 100.0%; Pred. No. 1.1e+03; Length 1453;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy
1745 ACAAGTCGACGCCAGCTGCC 1763
737 ACAAGTCGACGCCAGCTGCC 719

RESULT 6
AY418586
LOCUS
DEFINITION
AY418586 2013 bp DNA linear GSS 12-DEC-2003
Mus musculus HCM606 gene, VIRTUAL TRANSCRIPT, partial sequence.
ACCESSION
AY418586
VERSION
AY418586.1 GI:39774546
KEYWORDS
GSS.
SOURCE
Mus musculus (house mouse)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
Clark, A.G., Ghanowski, S., Nielson, R., Thomas, P., Kejarival, A.,
Todd, M.A., Tandenbaum, D.M., Civallo, D.R., Lu, F., Murphy, B.,
Ferriter, S., Wang, G., Zheng, X.H., White, T.J., Sninsky, J.J.,

```

TITLE
Adams,M.D. and Cargill,M.
Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL
Science 302 (5652), 1960-1963 (2003)
PUBMED
14671302
REFERENCE
2 (bases 1 to 2013)
AUTHORS
Clark,A.G., Glanowski,S., Nielsen,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civeille,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shtinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE
Direct Submission
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT
These sequences were made by sequencing genomic exons and ordering
them based on alignment
FEATURES
Location/Qualifiers
source
1..2013
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1..>2013
/locus_tag="HGM6606"
gene
ORIGIN
Query Match 1.0%; Score 19; DB 29; Length 2013;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1077 CGTGCCAGCGCTGCCGAGC 1095
|||||
Db 1132 CGTGCCAGCGCTGCCGAGC 1150
|||||
RESULT 7 2207 bp mRNA linear HTC 01-OCT-2002
LOCUS BC038231
DEFINITION Homo. sapiens, Similar to dual specificity phosphatase 8, clone
IMAGE:5547764, mRNA.
ACCESSION BC038231
VERSION BC038231.1 GI:23398534
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 2207)
AUTHORS Strausberg,R.
JOURNAL Direct Submission
PUBMED Submitted (30-SEP-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
COMMENT
NIH-MGC Project URL: http://mgc.ncl.nih.gov
Contact: MGC help desk
Email: cgaps-remail.nih.gov
Tissue Procurement: ATCC/DCTD/DTF
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gatherburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc.mgc@nih.gov
Akhter,N., Ayale,K., Beckstrom-Sternberg,S.M., Benjamin,B.,
Blakesley,R.W., Boufard,G.G., Breen,K., Brinkley,C., Brooks,S.,
Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghighi,P.,
Hansen,N., Ho,S.-L., Karlins,E., Kwong,P., Laric,P., Legardi,R.,
Maduro,Q.L., Masello,C., Maskeri,B., Magrini,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Stancipop,S., Thomas,P.J., Touchman,J.W.,
Taurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L.,
Young,A., Zhang,L.-H. and Green,E.D.
Clone distribution: MGC clone distribution information can be found

through the I.M.A.G.E. Consortium/LIML at: http://image.llnl.gov
Series: IRAK Plate: 79 Row: a Column: 12
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 4758211
This clone has the following problem: frame shifted.
FEATURES
Location/Qualifiers
source
1..2207
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5547764"
/issue_type="Skin, melanotic melanoma."
/clone_1ib="NIH MGC_72"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
ORIGIN
Query Match 1.0%; Score 19; DB 11; Length 2207;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 766 ATGCCTACATCATGAGA 784
|||||
Db 903 ATGCCTACATCATGAGA 921
|||||
RESULT 8 2289 bp DNA linear GSS 12-DEC-2003
LOCUS AY410969/c
DEFINITION Mus musculus HGM4067 gene, VIRTUAL TRANSCRIPT, partial sequence,
genomic survey sequence.
ACCESSION AY410969
VERSION AY410969.1 GI:39766937
KEYWORDS
SOURCE GSS.
ORGANISM Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 2289)
AUTHORS Clark,A.G., Glanowski,S., Nielsen,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civeille,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shtinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE
Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL
Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE
2 (bases 1 to 2289)
AUTHORS Clark,A.G., Glanowski,S., Nielsen,R., Thomas,P., Kejariwal,A.,
Todd,M.A., Tanenbaum,D.M., Civeille,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Shtinsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE
Direct Submission
Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT
This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
Location/Qualifiers
source
1..2289
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1..>2289
/locus_tag="HGM4067"
gene
ORIGIN
Query Match 1.0%; Score 19; DB 29; Length 2289;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1792 CGCAGCGCGCAGAGCCAA 1810
|||||
Db 246 CGCAGCGCGCAGAGCCAA 228
|||||

RESULT 9
AKO14482/c
LOCUS 2391 bp mRNA linear HTC 20-SEP-2003
DEFINITION Mus musculus 14 days embryo liver cDNA, RIKEN full-length enriched library, clone:4432409M07 product:hypothetical dDENN domain/DENN (AEX-3) domain containing protein, full insert sequence.
ACCESSION AKO14482
VERSION AKO14482.1 GI:12852366
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636

REFERENCE
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159

REFERENCE
AUTHORS Carninci, P., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sami, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861

REFERENCE
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE 5
AUTHORS The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
REFERENCE 6 (bases 1 to 2391)
AUTHORS Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H., Arawaka, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hirooka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Konno, H., Konda, M., Koyu, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Nunazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: genome-res@gs.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,

COMMENT

Fax: 81-45-503-9216)
Please visit our web site (<http://genome.gsc.riken.go.jp/>) for further details.

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer 15' GAGGAGCAGAGGATCCAGAGCTCTTTTCTTTTCTTTTAA 3', cDNA was prepared by using trichloroethanol-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence 15' GAGGAGCAGATTCGCGTTAATTATTAATCCCTCCCCCCCC 3'. cDNA was cleaved with BamHI and XhoI. Vector: a modified pBluescript KS(+) after bulk excision from lambdaBla FLIC 1. Cloning sites, 5' end: SalI, 3' end: BamHI. Host: DH10B.

FEATURES

source

1..2391
location/Qualifiers
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM DB:4432409M07"
/db_xref="MG1:1903235"
/db_xref="taxon:10090"
/clone="4432409M07"
/tissue_type="liver"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="14 days embryo"
110..913
/note="unnamed protein product; hypothetical dDENN domain/DENN (AEX-3) domain containing protein (Pfam|PF03455, InterPro|IPR01194, evidence: InterPro)
putative"
/codon_start=1
/protein_id="BAB29384.1"
/db_xref="GI:12852367"
/translation="MGSTRRRPRAPDMFEACGPNLSLEDPIIRKQPPDQBOEA
MOKVPRCFPPDIERPPSPAVQHTFTALTDLVGNRRFGCRLAGARSLCTLSHP
WFEVFKILNNVGDILAQNOVAEAEELNMLQOHPPLGPRSGSESDSITVRSBG
ILPPLAGNSKILSCFVAPDAASLPSIDENMLTELVAVDENIVGIFALMERVL
LTASKLSTVARDHRSRGDRITKPPFCVPLPSEGBEVAQWLFIGSTI"

CDS

ORIGIN

Query Match 1.0%; Score 19; DB 11; Length 2391;
Best Local Similarity 100.0%; Pred. No. 1.2e+03;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1792 CGCAGCGCGCAGAGCCAA 1810
|||||
Db 372 CGCAGCGCGCAGAGCCAA 354

RESULT 10
AKO40537/c
LOCUS 2985 bp mRNA linear HTC 19-SEP-2003
DEFINITION Mus musculus 0 day neonate thymus cDNA, RIKEN full-length enriched library, clone:4430105K15 product:hypothetical dDENN domain/DENN (AEX-3) domain containing protein, full insert sequence.

ACCESSION AKO40537
VERSION AKO40537.1 GI:26087915
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636

REFERENCE	2	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komoto, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. <i>Genome Res.</i> 10 (10), 1617-1630 (2000)
JOURNAL		
REFERENCE	3	Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Komoto, H., Akiyama, J., Nishi, K., Katsunuma, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, T., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer <i>Genome Res.</i> 10 (11), 1757-1771 (2000)
JOURNAL		
REFERENCE	4	20530913 11076861
JOURNAL		
REFERENCE	5	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. Functional annotation of a full-length mouse cDNA collection <i>Nature</i> 409, 685-690 (2001)
JOURNAL		
REFERENCE	6	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs <i>Nature</i> 420, 563-573 (2002)
JOURNAL		
REFERENCE	7	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fuyushida, K., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Hori, P., Imocani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawaji, J., Kojima, Y., Komoto, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Ohashi, N., Okazaki, Y., Saito, R., Saitoh, K., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takanashi, F., Takaku-Akahira, S., Takeda, Y., Tagami, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y. Direct Submission Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
JOURNAL		
REFERENCE	8	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL: http://genome.gsc.riken.go.jp/ URL: http://fantom.gsc.riken.go.jp/ Location/Qualifiers 1. 2985 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="FANTOM_DB:A430105K15" /db_xref="MGI:2405221" /db_xref="taxon:10090" /clone="A430105K15" /issue_type="thymus" /clone_type="RIKEN full-length enriched mouse cDNA library" /dev_stage="0 day neonate"
JOURNAL		
REFERENCE	9	
JOURNAL		
REFERENCE	10	
JOURNAL		
REFERENCE	11	
JOURNAL		
REFERENCE	12	
JOURNAL		
REFERENCE	13	
JOURNAL		
REFERENCE	14	
JOURNAL		
REFERENCE	15	
JOURNAL		
REFERENCE	16	
JOURNAL		
REFERENCE	17	
JOURNAL		
REFERENCE	18	
JOURNAL		
REFERENCE	19	
JOURNAL		
REFERENCE	20	
JOURNAL		
REFERENCE	21	
JOURNAL		
REFERENCE	22	
JOURNAL		
REFERENCE	23	
JOURNAL		
REFERENCE	24	
JOURNAL		
REFERENCE	25	
JOURNAL		
REFERENCE	26	
JOURNAL		
REFERENCE	27	
JOURNAL		
REFERENCE	28	
JOURNAL		
REFERENCE	29	
JOURNAL		
REFERENCE	30	
JOURNAL		
REFERENCE	31	
JOURNAL		
REFERENCE	32	
JOURNAL		
REFERENCE	33	
JOURNAL		
REFERENCE	34	
JOURNAL		
REFERENCE	35	
JOURNAL		
REFERENCE	36	
JOURNAL		
REFERENCE	37	
JOURNAL		
REFERENCE	38	
JOURNAL		
REFERENCE	39	
JOURNAL		
REFERENCE	40	
JOURNAL		
REFERENCE	41	
JOURNAL		
REFERENCE	42	
JOURNAL		
REFERENCE	43	
JOURNAL		
REFERENCE	44	
JOURNAL		
REFERENCE	45	
JOURNAL		
REFERENCE	46	
JOURNAL		
REFERENCE	47	
JOURNAL		
REFERENCE	48	
JOURNAL		
REFERENCE	49	
JOURNAL		
REFERENCE	50	
JOURNAL		
REFERENCE	51	
JOURNAL		
REFERENCE	52	
JOURNAL		
REFERENCE	53	
JOURNAL		
REFERENCE	54	
JOURNAL		
REFERENCE	55	
JOURNAL		
REFERENCE	56	
JOURNAL		
REFERENCE	57	
JOURNAL		
REFERENCE	58	
JOURNAL		
REFERENCE	59	
JOURNAL		
REFERENCE	60	
JOURNAL		
REFERENCE	61	
JOURNAL		
REFERENCE	62	
JOURNAL		
REFERENCE	63	
JOURNAL		
REFERENCE	64	
JOURNAL		
REFERENCE	65	
JOURNAL		
REFERENCE	66	
JOURNAL		
REFERENCE	67	
JOURNAL		
REFERENCE	68	
JOURNAL		
REFERENCE	69	
JOURNAL		
REFERENCE	70	
JOURNAL		
REFERENCE	71	
JOURNAL		
REFERENCE	72	
JOURNAL		
REFERENCE	73	
JOURNAL		

```

misc_feature      1..2985
                  /note="hypothetical dDENN domain/DENN (AEX-3) domain
                  containing protein (Pfam|PF03455, InterPro|IPR001194,
                  evidence: InterPro)"
ORIGIN
Query Match      1.0%; Score 19; DB 11; Length 2985;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY               1792 GCAGCGCGCAGAGCCAA 1810
                  |||||||
                  368 GCAGCGCGCAGAGCCAA 350

RESULT 11
LOCUS             BM926476               1447 bp      mRNA      linear      EST 12-MAR-2002
DEFINITION       AGNCNCURT_6644904 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5766863
                  5', mRNA sequence.
ACCESSION        BM926476
VERSION          BM926476.1
KEYWORDS         GI:19376855
SOURCE           EST.
ORGANISM         Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1447)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC Clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM12824 row: C column: 24
High quality sequence, stop: 557.
Location/Qualifiers
1..1447
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5766863"
/lab_host="DH10B"
/clone_1b="NIH_MGC_122"
/note="Organ: pooled lung and spleen; Vector: PCMV-SPORE;
Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
anonymous pool of 24 week female lung, 16 week female
spleen, and 20-22 week male spleens. Library is oligo-dT
primed and directionally cloned (EcoRV site is destroyed
upon cloning). Average insert size 1.4 kb, insert size
range 1-3 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(Invitrogen). Research Genetics tracking code 026. Note:
this is a NIH_MGC Library."
ORIGIN
Query Match      0.9%; Score 18; DB 12; Length 1447;
Best Local Similarity 100.0%; Pred. No. 3.5e+03;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY               819 TGTGAAGAAAAAGACC 836
                  |||||||
                  1236 TGTGAAGAAAAAGACC 1253

RESULT 12
RG18105/c

```

LOCUS BG118105 1506 bp mRNA linear EST 30-JAN-2001
 DEFINITION 602349869F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4444888 5',
 mRNA sequence.
 ACCESSION BG118105
 VERSION BG118105.1 GI:12611611
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1506)
 AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: ATCC
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LHAM10221 row: i column: 17
 High quality sequence stop: 2.
 Location/Qualifiers
 1..1506
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4444888"
 /tissue_type="adenocarcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH MGC_90"
 /note="Organ: liver; Vector: pCMV-SORT6; Site_1: NctI;
 Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 1.7 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."

ORIGIN

Query Match 0.9%; Score 18; DB 12; Length 1506;
 Best Local Similarity 100.0%; Pred. No. 3.5e+03;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1547 GCTTCCTTTCCGCGCTTT 1564
 |||||
 1506 GCTTCCTTTCCGCGCTTT 1489

RESULT 13
 CC206026 1545 bp DNA linear GSS 09-MAY-2003
 LOCUS CH261-191118_Sp6.1 CH261 Gallus gallus genomic clone CH261-191118,
 DEFINITION genomic survey sequence.
 ACCESSION CC206026
 VERSION CC206026.1 GI:30489567
 KEYWORDS GSS.
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianidae; Gallus.
 REFERENCE 1 (bases 1 to 1545)
 AUTHORS Krenitzki, C., Higginbotham, J., Wylie, K., Carter, J., McPherson, J.,
 Warren, W., Graves, T., Wards, E. and Wilson, R.
 TITLE Gallus gallus BAC End Reads
 JOURNAL Unpublished (2003)
 COMMENT Contact: Richard K. Wilson
 Genome Sequencing Center
 Washington University School of Medicine
 Email: submissions@wustl.edu
 Insert Length: 162000 Std Error: 0.00

Seq primer: Sp6 ATTTAGTCACACTATAG
 Class: BAC ends
 High quality sequence start: 250
 High quality sequence stop: 900.
 Location/Qualifiers
 1..1545
 /organism="Gallus gallus"
 /mol_type="genomic DNA"
 /strain="Red Jungle Fowl"
 /db_xref="taxon:9031"
 /clone="CH261-191118"
 /sex="female"
 /cell_line="UCD001, inbred 256"
 /clone_lib="CH261"
 /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
 CH261 Female Chicken Library - for library and clone
 ordering information: <http://www.chori.org/bacpac>"

ORIGIN

Query Match 0.9%; Score 18; DB 28; Length 1545;
 Best Local Similarity 100.0%; Pred. No. 3.5e+03;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 813 CAGATTGTGAAAGAAA 830
 |||||
 DB 816 CAGATTGTGAAAGAAA 799

RESULT 14
 BG757312 1564 bp mRNA linear EST 15-MAY-2001
 LOCUS 602715303F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4855296 5',
 DEFINITION mRNA sequence.
 ACCESSION BG757312
 VERSION BG757312.1 GI:14067965
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1564)
 AUTHORS NIH-MGC <http://mgs.nci.nih.gov/>.
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgabs-remail.nih.gov
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
 cDNA Library Preparation: Ling Hong/Rubin Laboratory
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
<http://image.llnl.gov>
 Plate: LNCM1704 row: f column: 01
 High quality sequence stop: 272.
 Location/Qualifiers
 1..1564
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:4855296"
 /tissue_type="primary B-cells from tonsils (cell line)"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NIH_MGC_48"
 /note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
 Site_2: EcoRI; cDNA made by oligo-dT priming.
 Directionally cloned into EcoRI/XhoI sites using the
 following 5' adaptor: GGACGAG(G). Size-selected >500bp
 for average insert size 1.8kb. Library constructed by Ling
 Hong in the laboratory of Gerald M. Rubin (University of
 California, Berkeley) using ZAP-cDNA synthesis kit
 (Stratagene) and Superscript II RT (Life Technologies).
 Note: this is a NIH_MGC Library."

ORIGIN

Query Match 0.9%; Score 18; DB 12; Length 1564;

Best Local Similarity 100.0%; Pred. No. 3.6e+03;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1690 TCACACTTCTACTCTGCC 1707
 |||||
 1276 TCACACTTCTACTCTGCC 1293

RESULT 15
 AK086250/c
 LOCUS
 DEFINITION AK086250 2647 bp mRNA linear HTC 20-SEP-2003
 Mus musculus 15 days embryo head cDNA, RIKEN full-length enriched
 library, clone:ID930016F01 product:unknown EST, full insert
 sequence.

ACCESSION AK086250
 VERSION AK086250.1 GI:26103338
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 AUTHORS Carninci, P. and Hayashizaki, Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636

2
 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Komoto, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
 JOURNAL Prepare full-length cDNA libraries for rapid discovery of new genes
 MEDLINE 20499374
 PUBMED 11042159

3
 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Komoto, H., Akiyama, J., Nishii, K., Katsunai, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
 Fujisake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
 TITLE RIKEN integrated sequence analysis (RISA) system—384-format
 JOURNAL sequencing pipeline with 384 multicapillary sequencer
 MEDLINE Genome Res. 10 (11), 1757-1771 (2000)
 PUBMED 20530913

4
 The RIKEN Genome Exploration Research Group Phase II Team and the
 JOURNAL PANTOM Consortium.
 MEDLINE Functional annotation of a full-length mouse cDNA collection
 PUBMED Nature 409, 685-690 (2001)

5
 The PANTOM Consortium and the RIKEN Genome Exploration Research
 JOURNAL Group Phase I & II Team.
 MEDLINE Analysis of the mouse transcriptome based on functional annotation
 PUBMED of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)

6
 (bases 1 to 2647)
 Adachi, J., Aizawa, K., Akimura, T., Arikawa, T., Bono, H., Carninci, P.,
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoaka, T., Hirozane, T.,
 Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
 Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Komoto, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishii, K., Nomura, K., Numazaki, R., Ohno, M., Ohnato, N.,
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,

Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
 Muramatsu, M. and Hayashizaki, Y.
 TITLE Direct Submission
 JOURNAL Submitted (16-APR-2002) Yoshinide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute; 1-7-22 Suenro-cho, Teurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)

COMMENT
 cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.go.jp/
 URL: http://fantom.gsc.riken.go.jp/
 Location/Qualifiers

FEATURES
 SOURCE
 1.2647
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="PANTOM:DB:D930016F01"
 /db_xref="MGI:2423638"
 /db_xref="taxon:10090"
 /clone="ID930016F01"
 /tissue_type="head"
 /clone_1ib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="15 days embryo"
 1..2647
 /note="unknown EST (GB|BB749563, evidence: BLASTN, 99%,
 match=375)"

misc_feature
 1..2647
 /note="unknown EST (GB|BB749563, evidence: BLASTN, 99%,
 match=375)"

ORIGIN
 Query Match 0.9%; Score 18; DB 11; Length 2647;
 Best Local Similarity 100.0%; Pred. No. 4.2e+03;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1465 ACCAGCAGCAGTGCGACC 1482
 |||||
 Db 2281 ACCAGCAGCAGTGCGACC 2264

Search completed: June 22, 2004, 09:21:26
 Job time : 7680 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 21, 2004, 18:35:51 ; Search time 1140 Seconds
(without alignments)
7434.347 Million cell updates/sec

Title: US-10-029-345A-108_COPY_538_2532

Perfect score: 1995
Sequence: 1 atggcccatgagatgattg9.....tggaaatcatgaggtctcc 1995

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 2124099041 residues

Word size : 0

Total number of hits satisfying chosen parameters: 772728

Minimum DB seq length: 1419

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_29Jun04:*

1: geneseq119808:.*
2: geneseq119808:.*
3: geneseq120008:.*
4: geneseq12001as:.*
5: geneseq12001as:.*
6: geneseq12002as:.*
7: geneseq12003as:.*
8: geneseq12003bs:.*
9: geneseq12003cs:.*
10: geneseq12004s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1995	100.0	5450	6	ACC60559 Polynucle
2	1995	100.0	5450	6	ACC60572 Polynucle
3	1893	94.9	3059	6	AAS15768 CDNA sequ
4	1893	94.9	3496	6	ABK47596 CDNA enco
5	1893	94.9	3544	5	AAS14639 Human CDN
6	1893	94.9	3544	6	ABK49402 CDNA enco
7	1893	94.9	3766	6	ABK14474 Human pro
8	1893	94.9	4790	6	ABN83966 Human gen
9	1893	94.9	5145	5	ABV20833 Human pro
10	1893	94.9	5145	5	ABV21080 Human pro
11	1893	94.9	5145	5	ABV26680 Human pro
12	1893	94.9	5145	5	ABV20978 Human pro
13	1893	94.9	5145	5	ABV21092 Human pro
14	1893	94.9	5145	5	ABV21312 Human pro
15	1893	94.9	5145	5	ABV21316 Human pro
16	1893	94.9	5145	5	ABV26826 Human pro
17	1893	94.9	5145	5	ABV27131 Human pro
18	1893	94.9	5145	5	ABV26923 Human pro
19	1893	94.9	5145	5	ABV27135 Human pro
20	1893	94.9	5145	5	ABV28657 Human pro
21	1893	94.9	5145	5	ABV22827 Human pro
22	1893	94.9	5145	5	ABV26934 Human pro
23	1849	92.7	2966	4	AAB99685 Human pro

24	1842	92.3	2732	4	AAD09492 Human SGP
25	1842	92.3	3104	6	ABN59704 Novel hum
26	1791	89.8	2102	7	ADA53105 Human cod
27	1791	89.8	5111	6	ACC60521 Polynucle
28	1593	79.8	2118	4	AAP30479 Human pro
29	1418	71.1	3332	6	ABK48378 CDNA enco
30	79	4.0	2756	6	ACC60560 Polynucle
31	21	1.1	2651	4	AAB54080 S. epider
32	21	1.1	4051	4	AAB54931 S. epider
33	20	1.0	1461	5	AAB72277 DNA enco
34	19	1.0	1422	6	ABL58978 HPV16-L2
35	19	1.0	1449	8	ADA02686 Human DUS
36	19	1.0	1449	9	ADB72424 Human DUS
37	19	1.0	1830	3	AAE64262 Human DUS
38	19	1.0	1909	6	ABT07757 Breathe ca
39	19	1.0	1909	8	ADA02685 Human DUS
40	19	1.0	1909	9	ADB72423 Human DUS
41	19	1.0	1917	9	ADE07547 Novel cod
42	19	1.0	2050	4	AAE63577 Human pho
43	19	1.0	2081	4	AAI99312 Human exc
44	19	1.0	2081	5	AAI63662 Human kid
45	19	1.0	2170	2	AAV52445 Streptoco

ALIGNMENTS

RESULT 1	ACC60559	standard; cdna, 5450 BP.
ID	ACC60559	
XX	ACC60559	
XX	19-JUN-2003 (first entry)	
XX	Polynucleotide relating to the invention SEQ ID NO: 108.	
XX	Gene, ser, antiproliferative; hepatotropic; nephrotoxic; antiarthritic;	
XX	antiproliferative; cardiatic; cytostatic; gene therapy; liver disease;	
XX	proliferative disorder; renal failure; cardiovascular disorder;	
XX	immunological disorder; arthritis; psoriasis; congenital heart defect;	
XX	congestive heart failure; cancer; human; mouse; yeast; rat; fruitfly.	
OS	Homo sapiens.	
XX	MO200257460-A2.	
XX	25-JUL-2002.	
XX	20-DEC-2001; 2001MO-US050459.	
XX	20-DEC-2000; 2000US-0256868P.	
XX	30-MAR-2001; 2001US-0280186P.	
XX	01-MAY-2001; 2001US-0287735P.	
XX	05-JUN-2001; 2001US-0295848P.	
XX	25-JUN-2001; 2001US-0300465P.	
XX	(BRIM) BRISTOL-MYERS SQUIBB CO.	
XX	Jackson DG, Feder J, Nelson T, Muntier G, Ramanathan C, Lee L;	
XX	Stiemers N, Bol D, Schieven G, Finger J, Todderud CG, Bassolino D;	
XX	Kyrtsek S, Mcatee P, Suchard S, Banas D;	
XX	WPI; 2002-559721/64.	
XX	P-PDSB; ABR52381.	
XX	Novel polynucleotides encoding human phosphatase polypeptides useful in	
XX	the prevention or treatment of e.g. proliferative and cardiovascular	
XX	disorders.	
XX	Claim 1, Fig 13; 801pp; English.	
XX	The invention relates to a novel isolated nucleic acid comprising a	

CC polynucleotide having a nucleotide sequence selected from 40
CC polynucleotides fully defined in the specification. The polynucleotide of
CC the invention has antiproliferative, hepatotropic, nephrotropic,
CC antiretroviral, antiparasitic, cardiant, and cytostatic activity. The
CC polynucleotide may have a use in gene therapy. A polynucleotide or
CC polypeptide of the invention is useful for preventing, treating or
CC ameliorating a medical condition, e.g. a proliferative disorder. They are
CC also useful for treating e.g. liver disease, renal failure, immunological
CC disorders including arthritis and psoriasis, cardiovascular disorders
CC such as congenital heart defects and congestive heart failure, and
CC cancer. A method of the invention is useful for diagnosing a pathological
CC condition or susceptibility to a condition in a subject. The present
CC sequence is used in the exemplification of the invention
XX

Sequence 5450 BP; 1471 A; 1211 C; 1337 G; 1431 T; 0 U; 0 Other;

Query Match 100.0%; Score 1995; DB 6; Length 5450;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1995; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGCCCATGAGATGATGGAACCTCAATTGTTCTGAGAGTGTGGCTGCTGGA 60
DB 538 ATGCCCATGAGATGATGGAACCTCAATTGTTCTGAGAGTGTGGCTGCTGGA 597
QY 61 AGTGAACGAAAAAGTCTCTAATTGATAGCCGCAATTGTGAATCAATACATCC 120
DB 598 AGTGAACGAAAAAGTCTCTAATTGATAGCCGCAATTGTGAATCAATACATCC 657
QY 121 CACATTTTGAAGGACATTAATATCACTGCTCCAGGCTTATGAAGGAGGTTGCAACAG 180
DB 658 CACATTTTGAAGGACATTAATATCACTGCTCCAGGCTTATGAAGGAGGTTGCAACAG 717
QY 181 GACAAAGTGTAAATTAACAGAGCTCATCAGCATTCAGGAAACATAAGTTGACATTGAT 240
DB 718 GACAAAGTGTAAATTAACAGAGCTCATCAGCATTCAGGAAACATAAGTTGACATTGAT 777
QY 241 TGCGTAGAAGAGTGTAAATTAACAGAGCTCATCAGCATTCAGGAAACATAAGTTGACATTGAT 300
DB 778 TGCGTAGAAGAGTGTAAATTAACAGAGCTCATCAGCATTCAGGAAACATAAGTTGACATTGAT 837
QY 301 GACTGTTTCTCACTGACTTCTGGGTAACCTGAGAGAGGCTCAACTCTGTTCACTG 360
DB 838 GACTGTTTCTCACTGACTTCTGGGTAACCTGAGAGAGGCTCAACTCTGTTCACTG 897
QY 361 CTTGCAAGTGGGTTGCTGAGTTCTCTCGTTGTTCCCTGCTCTGTAAGGAAATCC 420
DB 898 CTTGCAAGTGGGTTGCTGAGTTCTCTCGTTGTTCCCTGCTCTGTAAGGAAATCC 957
QY 421 ACTCTAGTCCCTAAGCTGATTTCTCAGGCTTGTACCTGTTGCCAATTTGGGCCAAC 480
DB 958 ACTCTAGTCCCTAAGCTGATTTCTCAGGCTTGTACCTGTTGCCAATTTGGGCCAAC 1017
QY 481 CGAATTTCTCCAAATCTTATCTTGGCTGCAGCGAGATGCTCTCAACAGAGGCTGATA 540
DB 1018 CGAATTTCTCCAAATCTTATCTTGGCTGCAGCGAGATGCTCTCAACAGAGGCTGATA 1077
QY 541 CAGCAGAAATGGGATTTGGTATGTTAAATGCCAGCTTATCCTGTCCAAAGCTGACTTT 600
DB 1078 CAGCAGAAATGGGATTTGGTATGTTAAATGCCAGCTTATCCTGTCCAAAGCTGACTTT 1137
QY 601 ATCCCGAGATCTCATTTCTGCGTGAGCTGTAATGACAGCTTTTGTGAAGAAATTTTG 660
DB 1138 ATCCCGAGATCTCATTTCTGCGTGAGCTGTAATGACAGCTTTTGTGAAGAAATTTTG 1197
QY 661 CGGTGTTTGAACAATCAGTAGATTTTCAATTGAGAAAGCAAAAGCTCCAAATGAGATGTT 720
DB 1198 CGGTGTTTGAACAATCAGTAGATTTTCAATTGAGAAAGCAAAAGCTCCAAATGAGATGTT 1257
QY 721 CTAGTGAATGTTTGTGGGATCTCCGCTCCGACCACTGCTATGCTTACATCATG 780
DB 1258 CTAGTGAATGTTTGTGGGATCTCCGCTCCGACCACTGCTATGCTTACATCATG 1317
QY 781 AAGAGATGAGATGATCTTTAGATGAAGCTTACAGATTTGTGAAGAAAAAGACCTAAT 840

DB 1318 AAGAGATGAGATGATCTTTAGATGAAGCTTACAGATTTTGTGAAGAAAAAGACCTACT 1377
QY 841 ATATCTCCAACTTCAATTTTCTGGGCCAACTCTGAGCTATGAGAAAGATTAAGAAC 900
DB 1378 ATATCTCCAACTTCAATTTTCTGGGCCAACTCTGAGCTATGAGAAAGATTAAGAAC 1437
QY 901 CAGACTGAGCATCAGAGGCAAGAGCAAACTCAAGCTGTGCTGAGAGAGGCAAT 960
DB 1438 CAGACTGAGCATCAGAGGCAAGAGCAAACTCAAGCTGTGCTGAGAGAGGCAAT 1497
QY 961 GAACTGTCTCTGCTGTCTCAGAGGCTGAGCAAGAAAGCAGAGCCCTCAGTCCACC 1020
DB 1498 GAACTGTCTCTGCTGTCTCAGAGGCTGAGCAAGAAAGCAGAGCCCTCAGTCCACC 1557
QY 1021 TGTGCGCATCTGTCTACCTCAGAGGCAAGAGCAAAAGCCCTGAGCTCCGAGGCTG 1080
DB 1558 TGTGCGCATCTGTCTACCTCAGAGGCAAGAGCAAAAGCCCTGAGCTCCGAGGCTG 1617
QY 1081 CCAGGCTGCGCAGAGCTGCAAGCCGTGCTGTGAAGAGCAGCCGCTGTACAGGCTC 1140
DB 1618 CCAGGCTGCGCAGAGCTGCAAGCCGTGCTGTGAAGAGCAGCCGCTGTACAGGCTC 1677
QY 1141 AGTGGGCTGACCTGTCCGAGACAGAGCTGGAAGACCAATTAAGCTCAAGCTTCTTC 1200
DB 1678 AGTGGGCTGACCTGTCCGAGACAGAGCTGGAAGACCAATTAAGCTCAAGCTTCTTC 1737
QY 1201 TCTCTGATTAATCAATCAGTTTCAATTAATCAGCAGATAGGAGGATCTTACATGCTTC 1260
DB 1738 TCTCTGATTAATCAATCAGTTTCAATTAATCAGCAGATAGGAGGATCTTACATGCTTC 1797
QY 1261 TCTCTATCAAGAGATGCTTTGGAATACTACAACTTTCATCTGATGAGGACCAAC 1320
DB 1798 TCTCTATCAAGAGATGCTTTGGAATACTACAACTTTCATCTGATGAGGACCAAC 1857
QY 1321 AAGCTATGCGAGTCTCCCTGTTGAGGAACTATCGAGAGAGCTCCCGAAACAGAGCTT 1380
DB 1858 AAGCTATGCGAGTCTCCCTGTTGAGGAACTATCGAGAGAGCTCCCGAAACAGAGCTT 1917
QY 1381 GATTAAGAGAGAGGCAAGCTATCCCAAGAGCTGAGACCGCCAGGCTTACAGACCCAG 1977
DB 1918 GATTAAGAGAGAGGCAAGCTATCCCAAGAGCTGAGACCGCCAGGCTTACAGACCCAG 1977
QY 1441 AGCAAGGATTTGATTTGGTCTGAGAACTGAGAGAGGCAAGCTTCCCTTTTAA 1500
DB 1978 AGCAAGGATTTGATTTGGTCTGAGAACTGAGAGAGGCAAGCTTCCCTTTTAA 2037
QY 1501 TCTCAGCTGATCGAAGTGGAGCTGAGAGCAATTAACACAGACGATCTCTTTTGGC 1560
DB 2038 TCTCAGCTGATCGAAGTGGAGCTGAGAGCAATTAACACAGACGATCTCTTTTGGC 2097
QY 1561 CTTTCCACAGGCAAGCACTCAAGAGCTGCTGAGGCTTGAAGGCTGAGCA 1620
DB 2098 CTTTCCACAGGCAAGCACTCAAGAGCTGCTGAGGCTTGAAGGCTGAGCA 2157
QY 1621 TCGGATATCTTGGGCCCCCAAGCTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1680
DB 2158 TCGGATATCTTGGGCCCCCAAGCTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2217
QY 1681 ACAAGTCTCAGACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740
DB 2218 ACAAGTCTCAGACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2277
QY 1741 GCTTACAGCTGAGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1800
DB 2278 GCTTACAGCTGAGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2337
QY 1801 CAGAAACCAAGTACAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1860
DB 2338 CAGAAACCAAGTACAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2397
QY 1861 AAGCACTTTAAACGAGAGCTGCAAAATGGAATTTGAGAGAGCATGCTCAGAGAAC 1920

Db 2398 AACGAGTTTAAACGACAGAGCTGCCAATGAAATTGAGAGAGCATCATGTCAGAGAAC 2457
QY 1921 AGGCAACGGAGAGAGCTGGGAGAAAGTGGGAGTCAGCTGTTTCCGGGACACATGGA 1980
Db 2458 AGGCAACGGAGAGCTGGGAGAAAGTGGGAGTCAGCTGTTTCCGGGACACATGGA 2517
QY 1981 ATCATTGAGCTCTCC 1995
Db 2518 ATCATTGAGCTCTCC 2532

RESULT 2
ACC60572
ID ACC60572 standard; cDNA; 5450 BP.
XX ACC60572;
AC ACC60572;
XX
XX 19-JUN-2003 (first entry)
XX
XX Polynucleotide relating to the invention SEQ ID NO: 147.
XX
XX Gene; ss; antiproliferative; hepatotropic; nephrotropic; antiarthritic;
XX antiproliferative; cardiant; cytosolic; gene therapy; liver disease;
XX antiproliferative disorder; renal failure; cardiovascular disorder;
XX immunological disorder; arthritis; psoriasis; congenital heart defect;
XX congestive heart failure; cancer; human; mouse; rat; fruitfly.
XX
XX Homo sapiens.
XX
XX WO200257460-A2.
XX
XX 25-JUL-2002.
XX
XX 20-DEC-2001; 2001WO-US050459.
XX
XX 20-DEC-2000; 2000US-0256868P.
XX 30-MAR-2001; 2001US-0280186P.
XX 01-MAY-2001; 2001US-0287735P.
XX 05-JUN-2001; 2001US-0295848P.
XX 25-JUN-2001; 2001US-0300465P.
XX
XX (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
XX Jackson DG, Feder J, Nelson T, Minter G, Ramanathan C, Lee L;
XX Stiemers N, Bol D, Schieven G, Finger J, Todderud CG, Bassolino D;
XX Krystek S, Mcatee P, Suchard S, Banas D;
XX
XX WPI; 2002-599721/64.
XX
XX P-PSDB; ABR52407.
XX
XX
XX Novel polynucleotides encoding human phosphatase polypeptides useful in
XX the prevention or treatment of e.g. proliferative and cardiovascular
XX disorders.
XX
XX Example 7; Fig 19; 801bp; English.

The invention relates to a novel isolated nucleic acid comprising a
polynucleotide having a nucleotide sequence selected from 40
polynucleotides fully defined in the specification. The polynucleotide of
the invention has antiproliferative, hepatotropic, nephrotropic,
antiarthritic, antiproliferative, cardiant, and cytostatic activity. The
polynucleotide may have a use in gene therapy. A polynucleotide or
amplifying a medical condition, e.g. a proliferative disorder. They are
also useful for treating e.g. liver disease, renal failure, immunological
disorders including arthritis and psoriasis, cardiovascular disorders
such as congenital heart defects and congestive heart failure, and
cancer. A method of the invention is useful for diagnosing a pathological
condition or susceptibility to a condition in a subject. The present
sequence is used in the exemplification of the invention

Sequence 5450 BP; 1471 A; 1211 C; 1337 G; 1431 T; 0 U; 0 Other;

Query Match 100.0%; Score 1995; DB 6; Length 5450;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1995; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCCCATGAGATGATGAACTCAATTTGTTACTGAGAGGTTGGCTCTGCTGGA 60
Db 538 ATGGCCCATGAGATGATGAACTCAATTTGTTACTGAGAGGTTGGCTCTGCTGGA 597
QY 61 AGTGAACCGAAAAAGTGTCTGCTAAATTGATACCGGCTTTTGTGAATCAATATATC 120
Db 598 AGTGAACCGAAAAAGTGTCTGCTAAATTGATACCGGCTTTTGTGAATCAATATATC 657
QY 121 CACATTTTGAAGCCATTATATCAACTGCTCCAACTTATGAAAGGAGTTGCAACAG 180
Db 658 CACATTTTGAAGCCATTATATCAACTGCTCCAACTTATGAAAGGAGTTGCAACAG 717
QY 181 GACAAAGTTTATTAACAGAGCTCATCCAGATTCCAGCAAACTAAGGTTGAATTGAT 240
Db 718 GACAAAGTTTATTAACAGAGCTCATCCAGATTCCAGCAAACTAAGGTTGAATTGAT 777
QY 241 TGCAGTCAGAGAGTTGATTTACATCAAGAGTCCCAAGATGTTGCTCTCTCTTCA 300
Db 778 TGCAGTCAGAGAGTTGATTTACATCAAGAGTCCCAAGATGTTGCTCTCTCTTCA 837
QY 301 GACTGTTTCTGACTGTACTTCTGGGTAACTGGAGAAAGGTTCAACTCTGTTCACTG 360
Db 838 GACTGTTTCTGACTGTACTTCTGGGTAACTGGAGAAAGGTTCAACTCTGTTCACTG 897
QY 361 CTTCAGAGTGGTTGCTGAGTCTCTCGTTGTTCCCTGGGCTCTGTGAAGAAATTC 420
Db 898 CTTCAGAGTGGTTGCTGAGTCTCTCGTTGTTCCCTGGGCTCTGTGAAGAAATTC 957
QY 421 ACTTATGTCCTCACTGCTCACTTTCTCAGGCTTGCTTACTGTTGCCAATTTGGCCATC 480
Db 958 ACTTATGTCCTCACTGCTCACTTTCTCAGGCTTGCTTACTGTTGCCAATTTGGCCATC 1017
QY 481 CGAATTTCTCCAACTTTATATCTGGCTGCGCAGCGAGATGCTCTCAACAGAGCTGATA 540
Db 1018 CGAATTTCTCCAACTTTATATCTGGCTGCGCAGCGAGATGCTCTCAACAGAGCTGATA 1077
QY 541 CAGCAAAATGGAATGTTATGTTAAATGACAGCTATACCTGTCCAAAGCTGACTTT 600
Db 1078 CAGCAAAATGGAATGTTATGTTAAATGACAGCTATACCTGTCCAAAGCTGACTTT 1137
QY 601 ATCCCGAGATCTCATTTCTCGCTGCTGCTGCTGATGAATGAAGCTTTGTGAATAATTTTG 660
Db 1138 ATCCCGAGATCTCATTTCTCGCTGCTGCTGCTGATGAATGAAGCTTTGTGAATAATTTTG 1197
QY 661 CCGTGGTTGCAAAATCAGTAGATTTCATTGAGAAAGCAAAAGCCCTCAATGATGTT 720
Db 1198 CCGTGGTTGCAAAATCAGTAGATTTCATTGAGAAAGCAAAAGCCCTCAATGATGTT 1257
QY 721 CTATGCACTGTTTATGCTGGATCTCCGCTCCGCAACCATGCTATCCCTACATCATG 780
Db 1258 CTATGCACTGTTTATGCTGGATCTCCGCTCCGCAACCATGCTATCCCTACATCATG 1317
QY 781 AAGAGATGACATGCTTTTATGATGAAGCTTACAGATTTGTGAAGAAAAAGACTTACT 840
Db 1318 AAGAGATGACATGCTTTTATGATGAAGCTTACAGATTTGTGAAGAAAAAGACTTACT 1377
QY 841 AATATCTCCAAATTTCAATTTTCTGGGCAACCTCTGACTATGAGAAAGATTAAAGAC 900
Db 1378 AATATCTCCAAATTTCAATTTTCTGGGCAACCTCTGACTATGAGAAAGATTAAAGAC 1437
QY 901 CAGACTGAGCATCAGGCGCAAGAGCAAACTCAAGCTGCTCAGCTGAGAGCAACCAAT 960
Db 1438 CAGACTGAGCATCAGGCGCAAGAGCAAACTCAAGCTGCTCAGCTGAGAGCAACCAAT 1497
QY 961 GAACCTGTCCTGCTGCTCTCAGAGGCTGACGAAAGAGAGAGCCCTCAGTCAACCC 1020
Db 1498 GAACCTGTCCTGCTGCTCTCAGAGGCTGACGAAAGAGAGAGCCCTCAGTCAACCC 1557
QY 1021 TGTGCGAATCTGCTACTCAGAGGACAGAGCAAAAGCCGCTGATCCGCAAGCTG 1080

[illegible][illegible]

QY 121 CAAATTGGAGACCAATTAATCAACTGCTCCAGCTTATGAAGCGAAGTTGCAACAG 180
DB 247 CAAATTTGGAAACCAATTAATCAACTGCTCCAGCTTATGAAGCGAAGTTGCAACAG 306
QY 181 GACAAAGTGTAAATTAACAGAGCTCATCCAGCATTCAGCGAACAATAGTTGACATTGAT 240
DB 307 GACAAAGTGTAAATTAACAGAGCTCATCCAGCATTCAGCGAACAATAGTTGACATTGAT 366
QY 241 TGAAGTCAGAAAGTTGATTAAGATCAAAAGCTCCCAAGATGTTGCTCTCTCTTCA 300
DB 367 TGAAGTCAGAAAGTTGATTAAGATCAAAAGCTCCCAAGATGTTGCTCTCTCTTCA 426
QY 301 GACTGTTTCTCACTGACTCTGAGTAACTGAGAAAGCTTCAACTCTGTTCACTG 360
DB 427 GACTGTTTCTCACTGACTCTGAGTAACTGAGAAAGCTTCAACTCTGTTCACTG 486
QY 361 CTTCAGAGTGGGTTTGCTGAGTTCTCTGTTGTTTCCCTGACCTCTGTGAGAAATTC 420
DB 487 CTTCAGAGTGGGTTTGCTGAGTTCTCTGTTGTTTCCCTGACCTCTGTGAGAAATTC 546
QY 421 ACTCTAGTCCCTACCTGATTTCTCAAGCTTGTGCTTACCTGTGCAACATTTGGCCCAAC 480
DB 547 ACTCTAGTCCCTACCTGATTTCTCAAGCTTGTGCTTACCTGTGCAACATTTGGCCCAAC 606
QY 481 CGAATCTTCCCAATCTTTATCTTGCTGACAGAGATGCTCTCAACAAGAGCTGAT 540
DB 607 CGAATCTTCCCAATCTTTATCTTGCTGACAGAGATGCTCTCTCAACAAGAGCTGAT 666
QY 541 CAGCAGAAATGGAAATGGTTATGTTTAAATGCCAGCTATACCTGTCCAAAGCTGACTTT 600
DB 667 CAGCAGAAATGGAAATGGTTATGTTTAAATGCCAGCTATACCTGTCTCAAAAGCTGACTTT 726
QY 601 ATCCCGAGTCTATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660
DB 727 ATCCCGAGTCTATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 786
QY 661 CCGTGGTTGGCAAAATCAGTAGATTTCAATGAGAAAGCAAAAGCTCCCAATGAGATGTT 720
DB 787 CCGTGGTTGGCAAAATCAGTAGATTTCAATGAGAAAGCAAAAGCTCCCAATGAGATGTT 846
QY 721 CTAGTGCACTGTTAGCTGGGATCTCCGCTCCGCCACATCGCTATCGCTACATCATG 780
DB 847 CTAGTGCACTGTTAGCTGGGATCTCCGCTCCGCCACATCGCTATCGCTACATCATG 906
QY 781 ABAAGATGACATGCTCTTGAATGAAGCTTACAGATTTGTGAAGAAAGAAAGCTACT 840
DB 907 ABAAGATGACATGCTCTTGAATGAAGCTTACAGATTTGTGAAGAAAGAAAGCTACT 966
QY 841 ATATCTCCAAACTTCAATTTCTGGGCACTCTGGAATGAGAAAGATTTAAGAAC 900
DB 967 ATATCTCCAAACTTCAATTTCTGGGCACTCTGGAATGAGAAAGATTTAAGAAC 1026
QY 901 CAGACTGAGACATCAGGGCCAAAGAGCAAACTCAAGCTGTGCACTGAGAAAGCAAAAT 960
DB 1027 CAGACTGAGACATCAGGGCCAAAGAGCAAACTCAAGCTGTGCACTGAGAAAGCAAAAT 1086
QY 961 GAACCTGTCTCTGCTGTCTCAGAGGTTGACAGAAAGCAAGCCCTCACTCAACCC 1020
DB 1087 GAACCTGTCTCTGCTGTCTCAGAGGTTGACAGAAAGCAAGCCCTCACTCAACCC 1146
QY 1021 TGTGCGCACTCTGCTGTCTCAGAGGTTGACAGAAAGCAAGCCCTCACTCAACCC 1080
DB 1147 TGTGCGCACTCTGCTGTCTCAGAGGTTGACAGAAAGCAAGCCCTCACTCAACCC 1206
QY 1081 CCCAGCGTCCAGAGCTGACAGCGCTGCTGTTAGAGGACAGCCCGCTGTTACAGGCGCT 1140
DB 1207 CCCAGCGTCCAGAGCTGACAGCGCTGCTGTTAGAGGACAGCCCGCTGTTACAGGCGCT 1266
QY 1141 AGTGGGCTGCACTGTCTCCGACAGAGCTGAGAAAGCAAGCTTAAAGCTCAAGCTTCTTC 1200
DB 1267 AGTGGGCTGCACTGTCTCCGACAGAGCTGAGAAAGCAAGCTTAAAGCTCAAGCTTCTTC 1326
QY 1201 TCTCTGATATCAAAATCAGTTTCAATTCAGCAAGCTGACAGATCTTAACTAGCTTC 1260

DB 1327 TCTCTGATATCAAAATCAGTTTCAATTCAGCAGAGATGAGCACTTACATGCTTC 1386
QY 1261 TCTCTATCAAGAAAGCTTTGGAAATCTAACAACTTCCACTACTTCTGATGGAGCAAC 1320
DB 1387 TCTCTATCAAGAAAGCTTTGGAAATCTAACAACTTCCACTACTTCTGATGGAGCAAC 1446
QY 1321 AAGCTATGCACTTCCCTGCTGTTAGAGAACTATGAGAGCACTCCGAAACCACTGCT 1380
DB 1447 AAGCTATGCACTTCCCTGCTGTTAGAGAACTATGAGAGCACTCCGAAACCACTGCT 1506
QY 1381 GATTAAGAGAAAGCCAGCATCTCCCAAGAAAGCTGACAGCCGCAAGCTTACAGACCAAG 1440
DB 1507 GATTAAGAGAAAGCCAGCATCTCCCAAGAAAGCTGACAGCCGCAAGCTTACAGACCAAG 1566
QY 1441 AGCAAGCAATTTGATTTGCTGCAAGAACACAGACAGATGGGACCGCCAGAGCTTCTTTA 1500
DB 1567 AGCAAGCAATTTGATTTGCTGCAAGAACACAGACAGATGGGACCGCCAGAGCTTCTTTA 1626
QY 1501 TCTCCACCTGCAATGAGTGGGAGCGTGGAGCAATTAACACACAGCTTCTTTTCCG 1560
DB 1627 TCTCCACCTGCAATGAGTGGGAGCGTGGAGCAATTAACACACAGCTTCTTTTCCG 1686
QY 1561 CTTTCCACAGCCAGCAGACAGCACTCAAGAGTCTGCTGCTGCTGCTTAAAGGCTGGCAC 1620
DB 1687 CTTTCCACAGCCAGCAGACAGCACTCAAGAGTCTGCTGCTGCTTAAAGGCTGGCAC 1746
QY 1621 TCGGATATCTTGGCCCCCAGACCTCTACCTTCCCTGACAGCAGCTGATTTTGGC 1680
DB 1747 TCGGATATCTTGGCCCCCAGACCTCTACCTTCCCTGACAGCAGCTGATTTTGGC 1806
QY 1681 ACAGAGTCTCAACTTCTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740
DB 1807 ACAGAGTCTCAACTTCTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1866
QY 1741 GCTACAGCTGACAGCAGCTGCTCACTTGGAGAACCAAGTATTTCTGCGCAGCGG 1800
DB 1867 GCTACAGCTGACAGCAGCTGCTCACTTGGAGAACCAAGTATTTCTGCGCAGCGG 1926
QY 1801 CAGAACCAAGTACAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1860
DB 1927 CAGAACCAAGTACAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1986
QY 1861 AAGCAATTTAAACGCAAGCTGCGCAATGGAATTTGAGAGAGCATGTCAGAGAAC 1920
DB 1987 AAGCAATTTAAACGCAAGCTGCGCAATGGAATTTGAGAGAGCATGTCAGAGAAC 2046
QY 1921 AGGTCAAGGGAAGAGCTGGGAAAGTGGCAGTCACTAGCTTTTCCGGCAGCATGAA 1980
DB 2047 AGGTCAAGGGAAGAGCTGGGAAAGTGGCAGTCACTAGCTTTTCCGGCAGCATGAA 2106
QY 1981 ATCATTGAGTCTCC 1995
DB 2107 ATCATTGAGTCTCC 2121
RESULT 4
ID ABK47596 standard; cDNA; 3496 BP.
XX ABK47596;
XX
XX 02-UTL-2002 (first entry)
XX
XX cDNA encoding human dual-specificity phosphatase-3 (DSP-16) protein.
DE Human: dual-specificity phosphatase-3; DSP-16; autoimmune disease;
XX mitogen-activated protein kinase; MAP; Duchenne muscular dystrophy;
XX cancer; graft-versus-host disease; allergy; metabolic disease;
XX abnormal cell growth; abnormal cell proliferation; contact inhibition;
XX cell cycle abnormality; anchorage independent cell growth; apoptosis;
XX intercellular adhesion; DSP-16 modulator; chromosome 12p; gene; ss.
XX

OS	Homo sapiens.
XX	
FH	Key
FT	Location/Qualifiers
CD	562..2559
FT	/tag= a
FT	/product= "Human dual-specificity phosphatase-3 (DSP-16)
FT	protein"
XX	
XX	
PX	WO200226997-A2.
PN	
PD	04-APR.-2002.
PF	25-SEP.-2001; 2001WO-US030124.
PR	26-SEP.-2000; 2000US-0235487P.
PA	(CEPT-) CEPTYR INC.
PI	Luche RM, Wei B;
PP	MPL; 2002-315602/35.
DR	P-PsDB; AAU79156.
XX	
XX	New DSP-16 polypeptide, useful for identifying modulators of its
PT	activity, which can be used in the treatment of disorders such as
PT	Duchenne muscular dystrophy, or cancer.
PS	Claim 7; Fig 1; 87pp; English.
XX	
XX	The present invention relates to a new polypeptide, DSP-16, having a 665
CC	amino acid sequence, given in the specification, or a variant having at
CC	least 50 % identical residues, which retains the ability to
CC	dephosphorylate an activated mitogen-activated protein (MAP) kinase. The
CC	invention can be used for identifying agents which modulate DSP-16
CC	activity, for modulation of a proliferative response in a cell, survival
CC	of a cell, or differentiation of a cell. The cell displays contact
CC	inhibition of cell growth or anchorage independent growth and may display
CC	altered intercellular adhesion. The agent may modulate apoptosis, or the
CC	cell cycle. The identified modulators can be used to treat Duchenne
CC	muscular dystrophy, cancer, graft-versus-host disease, autoimmune
CC	diseases, allergies, metabolic diseases, abnormal cell growth, abnormal
CC	cell proliferation, and cell cycle abnormalities. The present nucleic
CC	acid sequence is that of the human dual-specificity phosphatase-3 (DSP-
CC	16) gene located on chromosome 12p. This sequence encodes the human DSP-
CC	16 protein of the invention
XX	
XX	
SQ	Sequence 3456 BP; 939 A; 854 C; 856 G; 847 T; 0 U; 0 Other;
Query Match	94.9%; Score 1893; DB 6; Length 3456;
Best Local Similarity	99.9%; Pred. No. 0;
Matches 1993; Conservative	0; Mismatches 2; Indels 0; Gaps 0
OY	1 ATGGGCCCATGAGTATTTGAACCTCAATTGTACTAGAGGTGTGTGCTGTCTGAA 60
Db	562 ATGGCCCCAAGAATGATTTGAACTCAAATTTGTTACTGAGAGGTGTGTGCTGTCTGAA 621
OY	61 AGTGAACCGAAAAGTCGTGCTTAATTGATAGCCGCCCTTTGTGAAATACAATCATCC 120
Db	622 AATGGAACGAAAAAGTGCTGCTTAATTGATAGCCGCCCATTTGTGAAATACAATCATCC 681
OY	121 CACATTTTGGAAAGCCATTATAATATCAATGCTCCTCAAGCTATATGAGGAAGTTTGGAACG 180
Db	682 CACATTTTGGAAAGCCATTATAATATCAATGCTCCTCAAGCTATATGAGGAAGTTTGGAACG 741
OY	181 GACAAAGGTTAATTACAGAGCTATCCAGCACTTAGCGAAACATTAAGTTGCATTTGAT 240
Db	742 GACAAAGGTTAATTACAGAGCTATCCAGCACTTAGCGAAACATTAAGTTGCATTTGAT 801
OY	241 TGCACTCAGAAGTTGTAGTTTACATCAAAAGCTCCCAAGATGTGGCTCTCTCTTCA 300
Db	802 TGCACTCAGAAGTTGTAGTTTACATCAAAAGCTCCCAAGATGTGGCTCTCTCTTCA 861
OY	301 GACGTGTTTCACTGTACTTCTGGGTAACTGAGAAAGACCTTCACTGTTCACCTG 360

Db	862	GACTGTTTCTCAGCTGTAACCTCTGGGTAACGTGAGAAAGAGCTTCAACTCTGTTCACCTG	922
Qy	361	CTTGCAAGTGGGATTGTCGAGATTCTCTCGATTGTTCCCTGGCCCTCTGTGAAGAAATTC	422
Db	922	CTTGCAAGTGGGATTGTCGAGATTCTCTCGATTGTTCCCTGGCCCTCTGTGAAGAAATTC	981
Qy	421	ACTCTAGTCCTTACCTGCAATTTCTCAGCCCTTGCTTACCTGTGCAACATTTGGGCCAAC	480
Db	982	ACTCTAGTCCTTACCTGCAATTTCTCAGCCCTTGCTTACCTGTGCAACATTTGGGCCAAC	104
Qy	481	CGAATTCCTCCCAATCTTTATCTTGGGTGCACGAGAGATGCTCTCAACAAAGAGCTGATA	540
Db	1042	CGAATTCCTCCCAATCTTTATCTTGGGTGCACGAGAGATGCTCTCAACAAAGAGCTGATA	110
Qy	541	CAGCAATGCGATTGGTTATGTGTTAAATGCAAGCTATACCTGTCAAAAGCTGACTTT	600
Db	1102	CAGCAATGCGATTGGTTATGTGTTAAATGCAAGCTATACCTGTCAAAAGCTGACTTT	116
Qy	601	ATCCCGAGCTCATTTTCCGCGGTGGCCCTGTGAATGACAGACTTTTGTGAGAAATTTTG	660
Db	1162	ATCCCGAGCTCATTTTCCGCGGTGGCCCTGTGAATGACAGACTTTTGTGAGAAATTTTG	122
Qy	661	CCGTGTGTGACAAATCAGTAGATTTTCATTGTGAAGAAAGCAAGCCCTCAATGATGTGT	720
Db	1222	CCGTGTGTGACAAATCAGTAGATTTTCATTGTGAAGAAAGCAAGCCCTCAATGATGTGT	128
Qy	721	CTAGTGCATGTTTACTGTGGATTTCTCCGCTCCGCCACCATGCTATGCTATCATCATG	780
Db	1282	CTAGTGCATGTTTACTGTGGATTTCTCCGCTCCGCCACCATGCTATGCTATCATCATG	134
Qy	781	AAGAGATGACATGTCTTTAGATGAGCTTACAGATTTGTGAAGAAAGAAAGCACTAT	840
Db	1342	AAGAGATGACATGTCTTTAGATGAGCTTACAGATTTGTGAAGAAAGAAAGCACTAT	140
Qy	841	ATATCTCCCAACTCTTAATTTTGTGGGCCAACTCTTGGACATAGAGAAAGATTAAAGAC	900
Db	1402	ATATCTCCCAACTCTTAATTTTGTGGGCCAACTCTTGGACATAGAGAAAGATTAAAGAC	146
Qy	901	CAGACTGAGCATCAGGGCCAAAGAGCAAACTCAAGCTGTCGACCTTGGAGAGGCAAT	960
Db	1462	CAGACTGAGCATCAGGGCCAAAGAGCAAACTCAAGCTGTCGACCTTGGAGAGGCAAT	1521
Qy	961	GAACTGTTCCTCTGCTCTCAGAGGGTGGACAGAAAGCGAGACGCCCTCAGTCAACC	1020
Db	1522	GAACTGTTCCTCTGCTCTCAGAGGGTGGACAGAAAGCGAGACGCCCTCAGTCAACC	1581
Qy	1021	TGTGCGCACTCTGCTACCTCAGAGGCGACAGACAAAGGCCCGTGCATCCGCCACGGTG	1080
Db	1582	TGTGCGCACTCTGCTACCTCAGAGGCGACAGACAAAGGCCCGTGCATCCGCCACGGTG	1641
Qy	1081	CCGACGGTCCCAAGCGTGCAGCCGTCGTGTTAGAGACAGCCCGCTGTATCAGAGCGCTC	1140
Db	1642	CCGACGGTCCCAAGCGTGCAGCCGTCGTGTTAGAGACAGCCCGCTGTATCAGAGCGCTC	1701
Qy	1141	AGTGGGCTGCACCTGTCCGAGACAGAGGCTGGAACAGCAATTAAGTCAAGCGTTCCTTG	1200
Db	1702	AGTGGGCTGCACCTGTCCGAGACAGAGGCTGGAACAGCAATTAAGTCAAGCGTTCCTTG	1761
Qy	1201	TCTCTGATATCAAAATCAGTTTCATATTCAGCCAGCATGTCAGAGATCCTTACATGGCTTC	1260
Db	1762	TCTCTGATATCAAAATCAGTTTCATATTCAGCCAGCATGTCAGAGATCCTTACATGGCTTC	1821
Qy	1261	TCCCTCATCAAAAGTCTTTGGAAATACATAAACTTTCACATCTCTGTGATGGGACCAAC	1320
Db	1822	TCCCTCATCAAAAGTCTTTGGAAATACATAAACTTTCACATCTCTGTGATGGGACCAAC	1881
Qy	1321	AAGCTATGCCAGTTTCCCTGTTCAGAGAACTATGCGAGAGAGCTGCCAAACAGAGCTCT	1380
Db	1882	AAGCTATGCCAGTTTCCCTGTTCAGAGAACTATGCGAGAGAGCTGCCAAACAGAGCTCT	1941
Qy	1381	GATTAAGAGAAACGACATCCCAAGAAAGCTGACAGCCGCCAGGCTTTCAGACAGCCAG	1440

Db	1942	GATAAGGAGGAAGCCAGCATCCCAAGAAAGCTGGAGACCGCAGGACCTTCAGACAGCCAG	2001
Qy	1441	AGCAAGCATTGCAATTCGCTCAGAAACCGACAGCACTGACACCCGCCAGAGCTCCCTTTTA	1500
Db	2002	AGCAAGCATTGCAATTCGCTCAGAAACCGACAGCACTGACACCCGCCAGAGCTCCCTTTTA	2061
Qy	1501	TCTCCACTGCATTCGAAAGTGGAGGCGTGAAGGCAATTACCAACACAGCTTCCTTTTGCGC	1560
Db	2062	TCTCCACTGCATTCGAAAGTGGAGGCGTGAAGGCAATTACCAACACAGCTTCCTTTTGCGC	2121
Qy	1561	CTTTCACACAGCCAGCAGCAGCACTTCAGAAAGTCTGTGGCCCTGTGGCTTTAAGGCGCTGCGAC	1620
Db	2122	CTTTCACACAGCCAGCAGCAGCACTTCAGAAAGTCTGTGGCCCTGTGGCTTTAAGGCGCTGCGAC	2181
Qy	1621	TCGGATATCTTTGGCCCCCAGAACTCTAACCCCTTCCTGACCCAGACGCTGGTATTTTGGC	1680
Db	2182	TCGGATATCTTTGGCCCCCAGAACTCTAACCCCTTCCTGACCCAGACGCTGGTATTTTGGC	2241
Qy	1681	ACAGAGTCTCTACACTTACTCTGCGCCACGCACTACGAGAGGACAGTCCAGTACTCT	1740
Db	2242	ACAGAGTCTCTACACTTACTCTGCGCCACGCACTACGAGAGGACAGTCCAGTACTCT	2301
Qy	1741	GCCTACAGCTGACGCGCAGCTGCCCACTTGGCGAAGCAAGTCTATTCTGTGCGAGCGG	1800
Db	2302	GCCTACAGCTGACGCGCAGCTGCCCACTTGGCGAAGCAAGTCTATTCTGTGCGAGCGG	2361
Qy	1801	CAGAAAGCCAACTGACAGAGCTGACTCGCGCGGAGCTGGCATGAAGAGAGCCCTTTGAA	1860
Db	2362	CAGAAAGCCAACTGACAGAGCTGACTCGCGCGGAGCTGGCATGAAGAGAGCCCTTTGAA	2421
Qy	1861	AAGCAGTTTAAACGAGAAAGCTGCCAAATGGAAATTTGAGAGAGCATATATGTCAGAGAC	1920
Db	2422	AAGCAGTTTAAACGAGAAAGCTGCCAAATGGAAATTTGAGAGAGCATATATGTCAGAGAC	2481
Qy	1921	AGGTACAGGGGAAAGAGCTGGGGAAAGTGGGACGTCACTTTTGGGGCAGCATGGAA	1980
Db	2482	AGGTACAGGGGAAAGAGCTGGGGAAAGTGGGACGTCACTTTTGGGGCAGCATGGAA	2541
Qy	1981	ATCATTTGAGGTCTCC 1995	
Db	2542	ATCATTTGAGGTCTCC 2556	
RESULT 5			
AAS14639			
ID	AAS14639	standard; cDNA; 3544 BP.	
XX	AAS14639;		
AC			
XX			
DT	18-DEC-2001	(first entry)	
XX			
DE		Human cDNA encoding dual specificity phosphatase 21117.	
XX			
KM	Human; ss; dual specificity phosphatase 21117; hepatotropic; cytotabatic;		
KM	hematopoietic disorder; autoimmune disorder; diabetes mellitus;		
KM	rheumatoid arthritis; multiple sclerosis; Crohn's disease;		
KM	liver disorder; erythroid associated disorder; haemolytic anaemia;		
KM	cellular proliferative; differentiative disorder; leukaemia;		
KM	acute myeloid leukaemia; carcinoma; sarcoma; metastatic cancer.		
XX			
OS	Homo sapiens.		
XX			
XX	Key	Location/Qualifiers	
FX	CDS	589..2586	
FT		/*tag=	
FT		a	
FT		/product= "Dual specificity phosphatase 21117"	
XX			
XX	WO200173059-A2.		
XX			
XX	04-OCT-2001.		
XX			
PF	23-MAR-2001; 2001WO-US009477.		
XX			

PR	24-MAR-2000; 2000US-0191858P.
XX	(MILL-) MILLENNIUM PHARM INC.
PA	Meyers RA;
XX	WPI; 2001-611635/70.
DR	P-PSDB; AAU09016.
XX	New human dual specificity polypeptides and nucleic acids for diagnosis
PT	of disease and treatment of e.g. liver disorders.
XX	Claim 1; Fig 1; 143pp; English.
PS	
XX	The invention relates to two novel human dual specificity phosphatases
CC	designated 21117 and 38692, the nucleic acids encoding them (including
CC	fragments, allelic variants, their complements or nucleic acids that
CC	hybridise to them) and antibodies raised against the proteins. The
CC	antibody is useful for detecting the presence of the polypeptide, and the
CC	nucleic acid fragments are useful for detecting the presence of the
CC	nucleic acid, for diagnosing disease. Agents (e.g. the antibody and
CC	antisense sequences) are useful for modulating the activity or expression
CC	of the polypeptide/nucleic acid, useful for the treatment of e.g.
CC	hematopoietic disorders, autoimmune disorders (e.g. diabetes mellitus,
CC	rheumatoid arthritis, multiple sclerosis, Crohn's disease and many others
CC	listed in the specification) liver disorders, erythroid associated
CC	disorders (e.g. haemolytic anaemia) cellular proliferative or
CC	differentiative disorders, leukemias (e.g. acute myeloid leukaemia),
CC	carcinomas, sarcoma, metastatic cancers. Agents modulating 21117 and
CC	38692 are also useful for modulating the proliferation, survival,
CC	migration or differentiation of a 38692 or 21117-expressing cell. The
CC	polypeptide and nucleic acids are useful for identifying modulating
CC	agents. The present sequence encodes the dual specificity phosphatase
CC	21117
XX	
XX	Sequence 3544 BP; 1000 A; 840 C; 822 G; 882 T; 0 U; 0 Other;
XX	
Query Match	94.9%; Score 1893; DB 5; Length 3544;
Best Local Similarity	99.9%; Pred. No. 0;
Matches 1993; Conservative	0; Mismatches 2; Indels 0; Gaps 0;
QY	1 ATGGCCCATGAGATGATTTGAACTCAATTTGTTACTGAGAGTTGGTGGCTCTGCTGGAA 60
DB	589 ATGGCCCATGAGATGATTTGAACTCAATTTGTTACTGAGAGTTGGTGGCTCTGCTGGAA 648
QY	61 AGTGAACGCAAAAAGTCTCTAATTGATAGCCGGCCATTGTGGAAATACATATCATCC 120
DB	649 AGTGAACGCAAAAAGTCTCTAATTGATAGCCGGCCATTGTGGAAATACATATCATCC 708
QY	121 CACATTTTGGAAAGCATTAATATCAATCGTCCCAAGCTTATGAAGCGAAGGTGCAACAG 180
DB	709 CACATTTTGGAAAGCATTAATATCAATCGTCCCAAGCTTATGAAGCGAAGGTGCAACAG 768
QY	181 GACAAAGTGTAAATTACAGAGCATCATCAGCAATTCAGGAAACATAAGTTGACATTGAT 240
DB	769 GACAAAGTGTAAATTACAGAGCATCATCAGCAATTCAGGAAACATAAGTTGACATTGAT 828
QY	241 TGCAGTCGAAGAGTTTGTAGTTTACGATCAAAAGCTCCCAAGTGTGGCTCTCTCTTCA 300
DB	829 TGCAGTCGAAGAGTTTGTAGTTTACGATCAAAAGCTCCCAAGTGTGGCTCTCTCTTCA 888
QY	301 GACGTTTTTTCACATGTAATCTTGGAGTAAATCGAGAAAGACCTTCAACTCTGTTCAACTG 360
DB	889 GACGTTTTTTCACATGTAATCTTGGAGTAAATCGAGAAAGACCTTCAACTCTGTTCAACTG 948
QY	361 CTTCAGAGTGGGTTTGTGAGTTCTCTGTTGTTTCCCTGGCTCTGTGANGAANAATCC 420
DB	949 CTTCAGAGTGGGTTTGTGAGTTCTCTGTTGTTTCCCTGGCTCTGTGANGAANAATCC 1008
QY	421 ACTGTAGCTCTACTCTGATTTCTCAGCGTTGATCACTGTGTGCAACATTTGGGCAACC 480
DB	1009 ACTGTAGCTCTACTCTGATTTCTCAGCGTTGATCACTGTGTGCAACATTTGGGCAACC 1066

QY	481	CGAATTCCTCCCAATCTTAACTCTGGCTGCACACAGAAATGTCCTCAACAGAGACTGATA	540
Db	1069	CGAATTCCTCCCAATCTTAACTCTGGCTGCACACAGAAATGTCCTCAACAGAGAGCTGATG	1128
QY	541	CAGCAGAAATGGAAATGGTTAATGTTAAATGCCAGCTATACCTGTCCAAAGCCTGACTTT	600
Db	1129	CACACAGAAATGGAAATGGTTAATGTTAAATGCCAGCAATACCTGTCCAAAGCCTGACTTT	1188
QY	601	ATCCCGAGCTCAATTTCCGCGGTGGCCGTGAAATGACAAGCTTTTGTGAAAAATTTTG	660
Db	1189	ATCCCGAGCTCAATTTCCGCGGTGGCCGTGAAATGACAAGCTTTTGTGAAAAATTTTG	1248
QY	661	CCGTGTTGGACAAATCAGTAGATTTCAATTGAGAAAGCAAAAGCCTCCAAATGATGTGT	720
Db	1249	CCGTGTTGGACAAATCAGTAGATTTCAATTGAGAAAGCAAAAGCCTCCAAATGATGTGT	1308
QY	721	CTAGTCACTGTTTACTGGGATCTCCGCTCCGCCACCATCGTATCGCTCAATCATYG	780
Db	1309	CTAGTCACTGTTTACTGGGATCTCCGCTCCGCCACCATCGTATCGCTCAATCATYG	1368
QY	781	AAGAGAGTAGACATGCTTTAGATGAAAGCTTACAGTTTGTGAAAGAAAAAGACCTACT	840
Db	1369	AAGAGAGTAGACATGCTTTAGATGAAAGCTTACAGTTTGTGAAAGAAAAAGACCTACT	1428
QY	841	ATATCTCCAACTTCAATTTTCTGGGCCAACTCCTGGACTATGAGAAAMAGATTAAAGAC	900
Db	1429	ATATCTCCAACTTCAATTTTCTGGGCCAACTCCTGGACTATGAGAAAMAGATTAAAGAC	1488
QY	901	CAGACTGGAGCATCAGGGGCMAAAGAGCAAACTCAAGCTGTGCACCTGGAGAAACCTCAAT	960
Db	1489	CAGACTGGAGCATCAGGGGCMAAAGAGCAAACTCAAGCTGTGCACCTGGAGAAACCTCAAT	1548
QY	961	GAACTGTCTCTGTCTCTCAGAGGGTGGACAGAAAAGCAGAGCCCTTCAGTCCACCC	1020
Db	1549	GAACTGTCTCTGTCTCTCAGAGGGTGGACAGAAAAGCAGAGCCCTTCAGTCCACCC	1608
QY	1021	TGTGCGCACTCTGCTACCTCAGAGGACACAGAAAGCCCTGTGCATCCCGCAGAGGTG	1080
Db	1609	TGTGCGCACTCTGCTACCTCAGAGGACACAGAAAGCCCTGTGCATCCCGCAGAGGTG	1668
QY	1081	CCGACCGTGCACCGGTGCAGCCGTGTGTTAAGGACAGCCCGCTGGTACAGAGCGCTC	1140
Db	1669	CCGACCGTGCACCGGTGCAGCCGTGTGTTAAGGACAGCCCGCTGGTACAGAGCGCTC	1728
QY	1141	AGTGGGCTGCACTCTGCCGAGACAGGTGTGAAACAGCAATTAAGCTCAAGCTTCTCTC	1200
Db	1729	AGTGGGCTGCACTCTGCCGAGACAGGTGTGAAACAGCAATTAAGCTCAAGCTTCTCTC	1788
QY	1201	TCTGTGATATCAATCAGTTTCATATTCAGCCGACGTGGAGAGATCTTACATATGGCTTC	1260
Db	1789	TCTGTGATATCAATCAGTTTCATATTCAGCCGACGTGGAGAGATCTTACATATGGCTTC	1848
QY	1261	TCTCATATAGAAATGCTTTGGAATCTACAAACCTTCCACTCTCTGTGATGGAGCAAC	1320
Db	1849	TCTCATATAGAAATGCTTTGGAATCTACAAACCTTCCACTCTCTGTGATGGAGCAAC	1908
QY	1321	AAGCTATGCGAGTCTCCCGTGTTCAGAGAACTATCCGAGAGAGACTCCGAAACCACTCCT	1380
Db	1909	AAGCTATGCGAGTCTCCCGTGTTCAGAGAACTATCCGAGAGAGACTCCGAAACCACTCCT	1968
QY	1381	GATTAAGAGAGAACCGACATATCCCAAGAACTGTGCAGCCGACGAGCTTTCAGACAGCCAG	1440
Db	1969	GATTAAGAGAGAGAACCGACATATCCCAAGAACTGTGCAGCCGACGAGCTTTCAGACAGCCAG	2028
QY	1441	AGCAAGGCAATTCATTCGCTCAGAACACAGAGAGAGTGGCACCCGCCACAGAGTCCCTTTTA	1500
Db	2029	AGCAAGGCAATTCATTCGCTCAGAACACAGAGAGAGTGGCACCCGCCACAGAGTCCCTTTTA	2088
QY	1501	TCTCCACTGCATCGAAGTGGAGACGTGAGAGCAAAATTAACAACAACAGCTTCTTTTCGGC	1560
Db	2089	TCTCCACTGCATCGAAGTGGAGACGTGAGAGCAAAATTAACAACAACAGCTTCTTTTCGGC	2148
QY	1561	CTTTCCACCAAGCAGACACTTCAGAAATCTGTGCTGGCCCTTAAGGCGTGGCAC	1620

Db	2149	CTTTCCACCGCAGCAGACCACTCAGCAGAGTCTGTGGCTTGAAGGGCTGGCAC	2208
Qy	1621	TCGATATCTTGGCCCCCAGACCTTACCCCTTCCCTGACACAGCAGCTGGTATTTTGGC	1680
Db	2209	TGGGATATCTTGGCCCCCAGACCTCTACCCCTTCCCTGACACAGCAGCTGGTATTTTGGC	2268
Qy	1681	ACAGAGTCTCACAACCTTACTCTGTGCTCAGCCATCTACGGAGGCACTGCCAGTTACTCT	1740
Db	2269	ACAGAGTCTCACAACCTTACTCTGTGCTCAGCCATCTACGGAGGCACTGCCAGTTACTCT	2328
Qy	1741	GCTTACAGCTCAGACGACGTGCCCACTTGGCGGAAACCAATCTTATTTGTGGCAGGCGG	1800
Db	2329	GCTTACAGCTCAGACGACGTGCCCACTTGGCGGAAACCAATCTTATTTGTGGCAGGCGG	2388
Qy	1801	CAGAAGCCAAATGACAGAGCTGACTCGCGGCGAGCTGGCATGAAAGAGATCCCTTTGAA	1860
Db	2389	CAGAAGCCAAATGACAGAGCTGACTCGCGGCGAGCTGGCATGAAAGAGATCCCTTTGAA	2448
Qy	1861	AAGCAGTTTAAACGCAGAAAGCTGCCAAATGGAATTTGGAGAGAGCATCATGTACAGAAAC	1920
Db	2449	AAGCAGTTTAAACGCAGAAAGCTGCCAAATGGAATTTGGAGAGAGCATCATGTACAGAAAC	2508
Qy	1921	AGGTACAGGGGAGAAGCTGGGGAAAAGTGGGACATGACTTAAGCTTTTCGGGACAGATGAA	1980
Db	2509	AGGTACAGGGGAGAAGCTGGGGAAAAGTGGGACATGACTTAAGCTTTTCGGGACAGATGAA	2568
Qy	1981	ATCATTTAGAGTCTCC	1995
Db	2569	ATCATTTAGAGTCTCC	2583

XX	RESULT 6
XX	ABK49402
ID	ABK49402 standard; cDNA; 3544 BP.
XX	
AC	ABK49402;
XX	
DT	02-JUL-2002 (first entry)
XX	
DE	cDNA encoding human dual specificity phosphatase 21117 protein.
XX	
KW	Human; dual specificity phosphatase 21117; erythroid-related disorder;
KW	haematopoietic-related disorder; leukaemia; autoimmune disease; anaemia;
KW	erythrocytosis; liver-related disorder; cancer; gene; ss.
XX	
OS	Homo sapiens.
XX	
FH	Location/Qualifiers
FT	589..2586
CDS	/tag= a
FT	/product= "Human dual specificity phosphatase 21117"
FT	/note= "Specifically claimed in claim 2"
XX	
PN	US2002034807-A1.
XX	
PD	21-MAR-2002.
XX	
PF	23-MAR-2001; 2001US-00816494.
XX	
PR	24-MAR-2000; 2000US-0191858P.
XX	
PA	(MEYE/) MEYERS R A.
XX	
PI	Meyers RA;
XX	
DR	WPI; 2002-351088/38.
DR	P-PSDB; AAU79929.
XX	
PT	New nucleic acids, designated 38692 and 21117, encoding dual specificity
PT	phosphatases for treating cell proliferation and differentiation
PT	disorders including hematopoietic and erythroid-related disorders and
PT	cancers.

XX Claim 2; Fig 1; 76pp; English.

CC The present invention relates to new nucleic acids designated 38692 and
CC 2117 encoding dual specificity phosphatase family members. The nucleic
CC acid, polypeptide encoded by it, and antibody specific for the
CC polypeptide may be used to diagnose and treat haematopoietic-related
CC disorders such as leukaemias and autoimmune diseases, erythroid-related
CC disorders such as anaemias and erythrocytosis, liver-related disorders,
CC and cancers, particularly of the breast, colon, adipose, prostate and
CC lung. The present nucleic acid sequence encodes the human dual
CC specificity phosphatase 2117 protein of the invention, as described
CC above

XX Sequence 3544 BP; 1000 A; 840 C; 822 G; 882 T; 0 U; 0 Other;

Query Match 94.9%; Score 1893; DB 6; Length 3544;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCCCATGAGTGAATGTAAGTCAATTTGTAAGAGGTTGGGCTCTGCTGAA 60
DB ATGGCCCATGAGTGAATGTAAGTCAATTTGTAAGAGGTTGGGCTCTGCTGAA 648
QY 61 AGTGAACGGAAAGTGTCTAATTGATAGCGGCATTGGGAATACATPACATCC 120
DB AGTGAACGGAAAGTGTGTCTAATTGATAGCGGCATTGGGAATACATPACATCC 708
QY 121 CACATTTTGGAGCCATTATATCACTGCTCCAGCTTATGAGCGAGTTGCAACAG 180
DB CACATTTTGGAGCCATTATATCACTGCTCCAGCTTATGAGCGAGTTGCAACAG 768
QY 181 GACAAAGTGTATTTACAGAGCTCATCAGCATTCAGCGAAATTAAGTTGACATTGAT 240
DB GACAAAGTGTATTTACAGAGCTCATCAGCATTCAGCGAAATTAAGTTGACATTGAT 828
QY 241 TGCAGTCAGAAAGTGTAGTTACGATCAAGCTCCCAAGATGTGCTCTCTCTCA 300
DB TGCAGTCAGAAAGTGTAGTTACGATCAAGCTCCCAAGATGTGCTCTCTCTCA 888
QY 301 GACTGTTTCTCACTGACTTCTGGGTAACCTGGAGAAAGCTTCACTCTGTTCACTG 360
DB GACTGTTTCTCACTGACTTCTGGGTAACCTGGAGAAAGCTTCACTCTGTTCACTG 948
QY 361 CTTCGAGTGGTGTGCTGAGTTCTCTGTTGTTTCCCTGGCCTCTGTGAAGAAATCC 420
DB CTTCGAGTGGTGTGCTGAGTTCTCTGTTGTTTCCCTGGCCTCTGTGAAGAAATCC 1008
QY 421 ACTCTAGTCCCTACCTGCAATTTCTCAGCCTTGTGCTTACCTGTGCAACATTTGGCCAAAC 480
DB ACTCTAGTCCCTACCTGCAATTTCTCAGCCTTGTGCTTACCTGTGCAACATTTGGCCAAAC 1068
QY 481 CGAATTTCTCCCAATCTTTATCTTGGCTGCAGCGAGATGTCTCAACAGAGCTGATA 540
DB CGAATTTCTCCCAATCTTTATCTTGGCTGCAGCGAGATGTCTCAACAGAGCTGATA 1128
QY 1069 CGAATTTCTCCCAATCTTTATCTTGGCTGCAGCGAGATGTCTCAACAGAGCTGATA 1128
DB CGAATTTCTCCCAATCTTTATCTTGGCTGCAGCGAGATGTCTCAACAGAGCTGATA 1188
QY 541 CAGCAGAAATGGATGTGTAATGTTAAATGCCAGCTATACCTGTCCAAAGCCTGACTTT 600
DB CAGCAGAAATGGATGTGTAATGTTAAATGCCAGCTATACCTGTCCAAAGCCTGACTTT 1188
QY 601 ATCCCGAGTCTATTTCCCTGCGTGGCCGTGAAGTGAAGCAGCTTTTGTGAAGAAATTTTG 660
DB ATCCCGAGTCTATTTCCCTGCGTGGCCGTGAAGTGAAGCAGCTTTTGTGAAGAAATTTTG 1248
QY 661 CCGTGGTTGGACAATCAGTAGATTTCATGAGAAACAAAGACCTTCAATGATGTGT 720
DB CCGTGGTTGGACAATCAGTAGATTTCATGAGAAACAAAGACCTTCAATGATGTGT 1308
QY 721 CTAGTGCATGTTTAACTGGAGATCTCCGCTCCGCAACATCGCTATCGCTTACATCAG 780
DB CTAGTGCATGTTTAACTGGAGATCTCCGCTCCGCAACATCGCTATCGCTTACATCAG 1368
QY 781 AAGAGATGACATGTCTTAAATGAAGCTTACGATTTGTGAAGAAAGAAAGACCTACT 840

DB 1369 AAGAGATGACATGTCTTAAATGAAGCTTACGATTTGTGAAGAAAGAAAGACCTACT 1428
QY 841 ATATCTCCAAATCTCAATTTTCTGGGCAACTCTCTGACTATAGAGAAAGATTAAAGAC 900
DB ATATCTCCAAATCTCAATTTTCTGGGCAACTCTCTGACTATAGAGAAAGATTAAAGAC 1488
QY 901 CAGCTGAGCATCAGGGCCAAAGAGCAAACTCAAGCTGTGCACTGGAGAACCAAT 960
DB CAGCTGAGCATCAGGGCCAAAGAGCAAACTCAAGCTGTGCACTGGAGAACCAAT 1548
QY 961 GAACCTGTCCCTGTCTCTCAGAGGCTGAGCAGAAAGAGAGAGCCCTCACTGCAACC 1020
DB GAACCTGTCCCTGTCTCTCAGAGGCTGAGCAGAAAGAGAGAGCCCTCACTGCAACC 1608
QY 1021 TGTGCGCACTGTCTAATCTGAGAGCAGCAGCAAAAGCCGCTGATCCCGCAGCGTG 1080
DB TGTGCGCACTGTCTAATCTGAGAGCAGCAGCAGCAAAAGCCGCTGATCCCGCAGCGTG 1668
QY 1081 CCAGGCTCCAGCGGTGAGCCGTGCTGTTAAGAGACAGCCGCTGTGACAGGGGCTC 1140
DB CCAGGCTCCAGCGGTGAGCCGTGCTGTTAAGAGACAGCCGCTGTGACAGGGGCTC 1728
QY 1141 AGTGGCTGCACTGTCCGCAAGAGCTGAAAGCAGCAATPAGCTCAAGCGTTCTTC 1200
DB AGTGGCTGCACTGTCCGCAAGAGCTGAAAGCAGCAATPAGCTCAAGCGTTCTTC 1788
QY 1201 TCTTGGAATATCAATCAATCAATTTATTCAGCCAGATGAGCATCTTACATGCTTC 1260
DB TCTTGGAATATCAATCAATCAATTTATTCAGCCAGATGAGCATCTTACATGCTTC 1848
QY 1261 TCTCTATCAGAAAGTCTTTGGAATCTAATAACCTTCACTATCTGTGATGGAGCAAC 1320
DB TCTCTATCAGAAAGTCTTTGGAATCTAATAACCTTCACTATCTGTGATGGAGCAAC 1908
QY 1321 AAGCTATGCAAGTCTCCCTGTTTCAAGAACTATCGAGCAGACTCCCGAAACCAAGTCT 1380
DB AAGCTATGCAAGTCTCCCTGTTTCAAGAACTATCGAGCAGACTCCCGAAACCAAGTCT 1968
QY 1381 GATTAAGAGAAAGCCAGCATCTCCCAAGAAAGCTGAGCAGCCGACCTTACAGCAGCAG 1440
DB GATTAAGAGAAAGCCAGCATCTCCCAAGAAAGCTGAGCAGCCGACCTTACAGCAGCAG 2028
QY 1441 AGCAAGCATGATGATTCGCTCAGAACACAGCAGAGTGCAGCCGCAAGAGTCCCTTTA 1500
DB AGCAAGCATGATGATTCGCTCAGAACACAGCAGAGTGCAGCCGCAAGAGTCCCTTTA 2088
QY 1501 TCTCACTGATGAGAGTGGAGCGTGAAGCAATTTACACACAGCAGCTTCTTTCGAG 1560
DB TCTCACTGATGAGAGTGGAGCGTGAAGCAATTTACACACAGCAGCTTCTTTCGAG 2148
QY 1561 CTTTCACAGCAGCAGCAGCAGCTCAGAAAGTCTGCTGGCTGGGCTTAAAGGCTGGCAC 1620
DB CTTTCACAGCAGCAGCAGCAGCTCAGAAAGTCTGCTGGCTGGGCTTAAAGGCTGGCAC 2208
QY 1621 TCGGATATCTTGGCCCCCAGAGACTTACCCCTTCCCTGACCGAGCAGCTGTATTTTGGC 1680
DB TCGGATATCTTGGCCCCCAGAGACTTACCCCTTCCCTGACCGAGCAGCTGTATTTTGGC 2268
QY 1681 ACAGAGTCCCACTTCTACTGCTCAGCCATCTAAGAGAGAGAGTGCAGTTACTCT 1740
DB ACAGAGTCCCACTTCTACTGCTCAGCCATCTAAGAGAGAGAGTGCAGTTACTCT 2328
QY 1741 GCCTACAGTGCAGCAGCTGCCACTTGCAGAGACCAAGTCTATTTCTGTGCGCAGGCG 1800
DB GCCTACAGTGCAGCAGCTGCCACTTGCAGAGACCAAGTCTATTTCTGTGCGCAGGCG 2388
QY 1801 CAGAACCCAAAGTACAGAGCTGACTGCGCGGAGAGTGGCATGAAAGAGCCCTTTGAA 1860
DB CAGAACCCAAAGTACAGAGCTGACTGCGCGGAGAGTGGCATGAAAGAGCCCTTTGAA 2448
QY 1861 AAGCAGTTTAAAGCAGAGCTGCGCAATGGAATTTGAGAGAGCATGTGAGAGAAC 1920

DB 2449 AAGCAGTTTAAACGACGAGACTGCCAATGAAATTTGGAGAGACATCATGTCAGAGAAC 2508
QY 1921 AGGTCAACGGGAAAGAGCTGGGGAAGTGGGCACTCACTAGCTTTTCGGGACAGCTGGAA 1980
DB 2509 AGGTCAACGGGAAAGAGCTGGGGAAGTGGGCACTCACTAGCTTTTCGGGACAGCTGGAA 2568
QY 1981 ATCATTGAGGTCTCC 1995
DB 2569 ATCATTGAGGTCTCC 2583
RESULT 7
ID ABK14474 standard; cDNA; 3766 BP.
XX ABK14474;
AC ABK14474;
XX
DT 08-MAY-2002 (first entry)
XX
DE Human protein phosphatase 7 (PP7) cDNA sequence.
XX
KW Human; Protein phosphatase 7; PP7; immune system disorder; AIDS;
KW acquired immunodeficiency syndrome; thymic hypoplasia; anaemia; asthma;
KW Crohn's disease; neurological disorder; epilepsy; Huntington's disease;
KW dementia; Parkinson's disease; developmental disorder; Down's syndrome;
KW cell proliferative disorder; cancer; adenocarcinoma; leukaemia; lymphoma;
KW melanoma; myeloma sarcoma; gene; ss.
XX
OS Homo sapiens.
XX
XX
FH Location/Qualifiers
FT 538..2535
FT CDS /tag= a
FT /product= "Protein_phosphatase_7_(PP7)"
PN MO200210363-A2.
XX
PD 07-FEB-2002.
XX
PF 26-JUL-2001; 2001WO-US023716.
XX
PR 28-JUL-2000; 2000US-0221679P.
PR 03-AUG-2000; 2000US-0223272P.
PR 10-AUG-2000; 2000US-0224309P.
PR 18-AUG-2000; 2000US-0226728P.
PR 30-AUG-2000; 2000US-0229254P.
PR 08-SEP-2000; 2000US-0231366P.
XX
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Tang YT, Elliott VS, Ramkumar J, Yao MG, Burford N, Wang YE;
PI Stewart EA, Gandhi AR, Patterson C, Lee EA, Hafalia AUA, Lu DM;
PI Tribouley CM, Griffin JA, Baughn MR, Yue H, Warren BA, Nguyen DB;
PI Wallia NK, Kearney L;
XX
XX WPI: 2002-188735/24.
DR P-PSDB; AAU75789.
XX
XX
PT New protein phosphatases, useful for diagnosing, treating or preventing
PT immune system disorders (e.g. Crohn's disease); neurological disorders
PT (e.g. Parkinson's disease), or cell proliferative disorders (e.g.
PT cancers).
XX
XX Claim 5; Page 114-115; 117pp; English.
XX
XX The present invention relates to a new polypeptide, a naturally occurring
XX amino acid sequence at least 95 % identical to it, a biologically active
XX fragment of it or an immunogenic fragment of it. The polypeptides,
XX polynucleotides, agonists and antagonists are useful for diagnosing,
XX treating or preventing disorders associated with aberrant expression of
XX protein phosphatases (PP), particularly immune system disorders e.g.
XX acquired immunodeficiency syndrome (AIDS), thymic hypoplasia, anaemia,
XX asthma or Crohn's disease, neurological disorders e.g. epilepsy.

CC Huntington's disease, dementia or Parkinson's disease, developmental
CC disorders e.g. Down's syndrome, or cell proliferative disorders e.g.
CC cancers including adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma
CC or sarcoma. The present nucleic acid sequence encodes human protein
CC phosphatase 7 (PP7) which is one of several human protein phosphatases
CC (AAU75783-AAU75792) of the invention
XX
SQ Sequence 3766 BP; 1011 A; 903 C; 918 G; 934 T; 0 U; 0 Other;
Query Match 94.9%; Score 1893; DB 6; Length 3766;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGGCCCATGAGATGATGGAACTCAAAATGTTTACTGAGAGGTGTGCTGTGGAA 60
DB 538 ATGGCCCATGAGATGATGGAACTCAAAATGTTTACTGAGAGGTGTGCTGTGGAA 597
QY 61 AGTGAACGGAAAAAGTGTGCTAATTGATAGCCGGCCATTGTGGAATTACAATCATCC 120
DB 598 AGTGAACGGAAAAAGTGTGCTAATTGATAGCCGGCCATTGTGGAATTACAATCATCC 657
QY 121 CACATTTTGAAGCCATTAAATATCAATGCTTCAAGCTTATGAAGCGAAGTTGCAACAG 180
DB 658 CACATTTTGAAGCCATTAAATATCAATGCTTCAAGCTTATGAAGCGAAGTTGCAACAG 717
QY 181 GACAAAGCTTAATTAACAGAGCTCATCCAGCATTCAGCGAAACATAGTTGAACATTGAT 240
DB 718 GACAAAGCTTAATTAACAGAGCTCATCCAGCATTCAGCGAAACATAGTTGAACATTGAT 777
QY 241 TGCAGTCAGAAAGTGTGATTTAGATCAAGAGCTCCCAAGATGTGCTCTCTCTTCA 300
DB 778 TGCAGTCAGAAAGTGTGATTTAGATCAAGAGCTCCCAAGATGTGCTCTCTCTTCA 837
QY 301 GACTGTTTCTCACTGTACTTCTGAGTAACTGAGAAAGAGCTTCAACTCTGTTACCTG 360
DB 838 GACTGTTTCTCACTGTACTTCTGAGTAACTGAGAAAGAGCTTCAACTCTGTTACCTG 897
QY 361 CTTCGAGTGGGTTTGCCTGAAATTCCTGCTGTTTCCCTGCGCTCTGTGAAGAAATTC 420
DB 898 CTTCGAGTGGGTTTGCCTGAAATTCCTGCTGTTTCCCTGCGCTCTGTGAAGAAATTC 957
QY 421 ACTTAGTCCCTACCTGATTTCTCAGGCTTCTTACCTGTGCAACATTTGGGCAAC 480
DB 958 ACTTAGTCCCTACCTGATTTCTCAGGCTTCTTACCTGTGCAACATTTGGGCAAC 1017
QY 481 CGAATTTCTCCCAATCTTATCTTGTGCTGCCAGGAGATGCTTCAACAGAGCTGATA 540
DB 1018 CGAATTTCTCCCAATCTTATCTTGTGCTGCCAGGAGATGCTTCAACAGAGCTGATG 1077
QY 541 CAGCAGATGGGATTTGTTATGTTTAAATGCGACGTAATACCTGTCAAAAGCTGACTTT 600
DB 1078 CAGCAGATGGGATTTGTTATGTTTAAATGCGACGTAATACCTGTGTCAAAAGCTGACTTT 1137
QY 601 ATCCCGAGATCTCAATTTCTGCGTGTGCTGTGAATGACAGCTTTTGTGAGAAATTTTG 660
DB 1138 ATCCCGAGATCTCAATTTCTGCGTGTGCTGTGAATGACAGCTTTTGTGAGAAATTTTG 1197
QY 661 CCGTGTGTTGACAAATAGTAGATTTTATGAGAAAGCAAAAGCTTCAATGATGTGTT 720
DB 1198 CCGTGTGTTGACAAATAGTAGATTTTATGAGAAAGCAAAAGCTTCAATGATGTGTT 1257
QY 721 CTAGTGACATGTTTAACTGGGATCTCCGCTCGGCCACCAATGCTATGCTATCATCATG 780
DB 1258 CTAGTGACATGTTTAACTGGGATCTCCGCTCGGCCACCAATGCTATGCTATCATCATG 1317
QY 781 AAGAGATGACATGCTTTAGATGAAGCTTACAGATTTGTGAAAGAAAAAGACTTACT 840
DB 1318 AAGAGATGACATGCTTTAGATGAAGCTTACAGATTTGTGAAAGAAAAAGACTTACT 1377
QY 841 ATATCTCCAACTTCAATTTTCTGGGCCAACTCTGCACTATGAGAAAGATTAAGAAC 900
DB 1378 ATATCTCCAACTTCAATTTTCTGGGCCAACTCTGCACTATGAGAAAGATTAAGAAC 1437

PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

PI Schlegel R, Endege WO, Monahan JE.

DR WPI; 2001-662795/76.

XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.

XX Claim 1; Page 3419; 11750DP; English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
XX

SO Sequence 5145 BP; 1408 A; 1135 C; 1253 G; 1346 T; 0 U; 3 Other;

Query Match 94.9%; Score 1893; DB 5; Length 5145;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCCCAAGAGTGAATGGAAGTCAATTTGTTACTGAGAGTGGTGGCTGCGTGA 60
DB ATGGCCCAAGAGTGAATGGAAGTCAATTTGTTACTGAGAGTGGTGGCTGCGTGA 648
QY 61 AGTGAACGGAAGAGTGTCTAATGATAGCCGCGCATTTGTGATCAATATATATCC 120
DB AGTGAACGGAAGAGTGTCTAATGATAGCCGCGCATTTGTGATCAATATATATCC 708
QY 121 CACATTTTGGAGCCATTAATATCAACTGCTCCAGCTTATGAAGCGAAGTGGCAAC 180
DB CACATTTTGGAGCCATTAATATCAACTGCTCCAGCTTATGAAGCGAAGTGGCAAC 768
QY 181 GACAAAGTGTATATACAGAGCTCATTCAGCATTCAGCGAAACATTAAGTTGACATTGAT 240
DB GACAAAGTGTATATACAGAGCTCATTCAGCATTCAGCGAAACATTAAGTTGACATTGAT 828
QY 241 TGCAGTCAGAAAGTGTATGATGATCAAGAGTCCCAAGATGTGGCTCTCTCTTCA 300
DB TGCAGTCAGAAAGTGTATGATGATCAAGAGTCCCAAGATGTGGCTCTCTCTTCA 888
QY 301 GACTGTTTCTCACTGACTTCTGGGTAACCTGAGAAAGCTTCAACTCTGTTCACTG 360
DB GACTGTTTCTCACTGACTTCTGGGTAACCTGAGAAAGCTTCAACTCTGTTCACTG 948
QY 361 CTTCGAGTGGGTTTGTGAGTTCTCTGTTTTCCTTGGCTCTCTGTAAGAAATTC 420
DB CTTCGAGTGGGTTTGTGAGTTCTCTGTTTTCCTTGGCTCTCTGTAAGAAATTC 1008
QY 421 ACTCTAGTCCCTACCTGATTTCTCAGCTTGTGCTTATGCAATTTGGGCCAAC 480
DB ACTCTAGTCCCTACCTGATTTCTCAGCTTGTGCTTATGCAATTTGGGCCAAC 1068
QY 481 CGAATCTTCCCAATCTTATCTTGGCTGCGAGAGATGTCTCAACAAAGAGCTGATA 540
DB CGAATCTTCCCAATCTTATCTTGGCTGCGAGAGATGTCTCAACAAAGAGCTGATA 1128

QY 541 CAGCAGAAATGGATTTGTTATGTTAAATGCGAGCTATATCTGTCCAAAGCTGACTT 600
DB CAGCAGAAATGGATTTGTTATGTTAAATGCGAGCTATATCTGTGTCCAAAGCTGACTT 1188
QY 601 ATCCCGAGTCTCATTTCTGCGTGTGCTGTGATGACAGCTTTTGTGAAATTTTG 660
DB ATCCCGAGTCTCATTTCTGCGTGTGCTGTGATGACAGCTTTTGTGAAATTTTG 1248
QY 661 CCGTGTGGAGCAATACATTAATTTGATGAAAGCAAAAGCCCTCAATGGATGTT 720
DB CCGTGTGGAGCAATACATTAATTTGATGAAAGCAAAAGCCCTCAATGGATGTT 1308
QY 721 CTATGCACTGTTTATGCTGGATCTCCGCTCGGACCATGCTATGCTCTCATCATG 780
DB CTATGCACTGTTTATGCTGGATCTCCGCTCGGACCATGCTATGCTCTCATCATG 1368
QY 781 AAGAGATGAGATGCTTTTAAATGAGCTTACAGATTTGTGAAAGAAAGAAAGCTACT 840
DB AAGAGATGAGATGCTTTTAAATGAGCTTACAGATTTGTGAAAGAAAGAAAGCTACT 1428
QY 841 ATATCTCCAACTTCAATTTTCTGGGCAACTCTGCTGATATGAAAGAAAGTTAAGAC 900
DB ATATCTCCAACTTCAATTTTCTGGGCAACTCTGCTGATATGAAAGAAAGTTAAGAC 1488
QY 901 CAGACTGAGCATYACAGGCGCAAGCAAACTCAAGCTGCTGCACTGGAAGCCAAAT 960
DB CAGACTGAGCATYACAGGCGCAAGCAAACTCAAGCTGCTGCACTGGAAGCCAAAT 1548
QY 961 GAACTGTCTGCTGTCTCAAGAGGTGACAGAAAGCGAGAGCGCTTCACTGACACC 1020
DB GAACTGTCTGCTGTCTCAAGAGGTGACAGAAAGCGAGAGCGCTTCACTGACACC 1608
QY 1021 TGTGCGGATCTGTCTCAAGAGGTGACAGAAAGCGAGAGCGCTTCACTGACACC 1080
DB TGTGCGGATCTGTCTCAAGAGGTGACAGAAAGCGAGAGCGCTTCACTGACACC 1668
QY 1609 TGTGCGGATCTGTCTCAAGAGGTGACAGAAAGCGAGAGCGCTTCACTGACACC 1668
DB TGTGCGGATCTGTCTCAAGAGGTGACAGAAAGCGAGAGCGCTTCACTGACACC 1728
QY 1081 CCAGCGTCCAGCGTGTGAGCGCTGTGTTAGAGAGAGCGCGCTGTTAGAGAGCGCT 1140
DB CCAGCGTCCAGCGTGTGAGCGCTGTGTTAGAGAGAGCGCGCTGTTAGAGAGCGCT 1728
QY 1141 AGTGGCTGACCTGTCTGAGAGAGCGCTGTGTTAGAGAGAGCGCGCTGTTAGAGAGCGCT 1200
DB AGTGGCTGACCTGTCTGAGAGAGCGCTGTGTTAGAGAGAGCGCGCTGTTAGAGAGCGCT 1788
QY 1729 AGTGGCTGACCTGTCTGAGAGAGCGCTGTGTTAGAGAGAGCGCGCTGTTAGAGAGCGCT 1788
DB AGTGGCTGACCTGTCTGAGAGAGCGCTGTGTTAGAGAGAGCGCGCTGTTAGAGAGCGCT 1848
QY 1201 TCTCTGATATCAATCAATGTTTCTATTTAGAGAGAGCGCGCTGTTAGAGAGCGCT 1260
DB TCTCTGATATCAATCAATGTTTCTATTTAGAGAGAGCGCGCTGTTAGAGAGCGCT 1848
QY 1261 TCTCTGATATCAATCAATGTTTCTATTTAGAGAGAGCGCGCTGTTAGAGAGCGCT 1320
DB TCTCTGATATCAATCAATGTTTCTATTTAGAGAGAGCGCGCTGTTAGAGAGCGCT 1908
QY 1321 AAGCTATGCAATTTCTCCCTGTTTCAAGAACTATCGAGAGAGCTCCGAAACCAAGTCT 1380
DB AAGCTATGCAATTTCTCCCTGTTTCAAGAACTATCGAGAGAGCTCCGAAACCAAGTCT 1968
QY 1381 GATAGAGAGAGAGCGCATTCGCCAAGAGCTGCAACCGCGCGCTTCAAGAGAGCG 1440
DB GATAGAGAGAGAGCGCATTCGCCAAGAGCTGCAACCGCGCGCTTCAAGAGAGCG 2028
QY 1441 AGCAAGCAATGATGCTGCTGAGAGAGCGAGAGAGCGAGAGAGCTTCTTTT 1500
DB AGCAAGCAATGATGCTGCTGAGAGAGCGAGAGAGCGAGAGAGCTTCTTTT 2088
QY 1501 TCTTCACTGATGAG 1560
DB TCTTCACTGATGAG 2148
QY 1561 CTCTTCACTGATGAG 1620
DB CTCTTCACTGATGAG 2208
QY 1621 TCGGATATCTTGGCCCCCGAG 1680

Db 2209 TCGATATCTTGCCCCCAGACCTCTACCCCTTCCCTGACACAGCTGTATTTTCC 2268
Qy 1661 ACAGAGTCTCACTTCTACTCTGCTCAAGCCATCTTAACGAGGCGAGTGCAGTTACTCT 1740
Db 2269 ACAGAGTCTCACTTCTACTCTGCTCAAGCCATCTTAACGAGGCGAGTGCAGTTACTCT 2228
Qy 1741 GCCTACAGCTGACGACGCTGCTCCACTTGGCGAGACCAAGCTTATTTCTGTCGGAGCGG 1800
Db 2229 GCCTACAGCTGACGACGCTGCTCCACTTGGCGAGACCAAGCTTATTTCTGTCGGAGCGG 2288
Qy 1801 CAGAGCCAACTGACAGAGCTGACTCGCGCGAGCTGAGTGAAGAGAGCCCTTTGAA 1860
Db 2389 CAGAGCCAACTGACAGAGCTGACTCGCGCGAGCTGAGTGAAGAGAGCCCTTTGAA 2448
Qy 1861 AAGCAGTTTAAACCCAGAGAGCTGCGCAATGATTTGAGAGAGCATGTCAGAGAAC 1920
Db 2449 AAGCAGTTTAAACCCAGAGAGCTGCGCAATGATTTGAGAGAGCATGTCAGAGAAC 2508
Qy 1921 AGCTCAGCGGAAAGCTGGGGAAAGTGAGGAGTCACTAGCTTTTGGCGAGCATGGA 1980
Db 2509 AGCTCAGCGGAAAGCTGGGGAAAGTGAGGAGTCACTAGCTTTTGGCGAGCATGGA 2568
Qy 1981 ATCATTGAGGTCTCC 1995
Db 2569 ATCATTGAGGTCTCC 2583

RESULT 10

ABV21080
ID ABV21080 standard; cDNA; 5145 BP.

XX AC ABV21080;

DT 13-SEP-2002 (first entry)

DE Human prostate expression marker cDNA 21071.

KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;

XX KW pharmacogenomic marker; gene; ss.

OS Homo sapiens.

PN WO200160860-A2.

PD 23-AUG-2001.

PF 20-FEB-2001; 2001WO-US005171.

PR 17-FEB-2000; 2000US-018319P.

PR 16-MAR-2000; 2000US-0189862P.

PR 25-MAY-2000; 2000US-0207454P.

PR 09-JUN-2000; 2000US-0211314P.

PR 18-JUL-2000; 2000US-0219007P.

PR 13-DEC-2000; 2000US-0255281P.

PA (MIL-) MILLENNIUM PREDICTIVE MEDICINE INC.

PI Schlegel R, Endege WO, Monahan UE;

PT WPI; 2001-662795/76.

PT Novel isolated nucleic acid molecule associated with cancerous state of

CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the

CC specification of its complement. (I) is useful for: (a) assessing whether

CC a patient is afflicted with prostate cancer; (b) monitoring the

CC progression of prostate cancer in a patient; (c) assessing the efficacy

CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
SQ Sequence 5145 BP; 1408 A; 1135 C; 1253 G; 1346 T; 0 U; 3 Other;

Query Match 94.9%; Score 1993; DB 5; Length 5145;
Best Local Similarity 99.9%; Pred No 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGCCCATGAGATGATTTGAACTCAATTTGTTACTGAGAGGTGTGCTCTGCGAA 60
Db 589 ATGGCCCATGAGATGATTTGAACTCAATTTGTTACTGAGAGGTGTGCTCTGCGAA 648
Qy 61 AGTGAACGGAAGAAAGTGTCTGTAATTGATAGCCGCGCAATTTGGAATACATATCATCC 120
Db 649 AGTGAACGGAAGAAAGTGTCTGTAATTGATAGCCGCGCAATTTGGAATACATATCATCC 708
Qy 121 CACATTTTGAAGACCATTAATATCACTGCTCCAAAGTTATGAAAGCGAAGGTGCAAG 180
Db 709 CACATTTTGAAGACCATTAATATCACTGCTCCAAAGTTATGAAAGCGAAGGTGCAAG 768
Qy 181 GACAAAGTGTAAATTACAGAGCTCAATCCAGATTCAGGAAACATTAAGTTGACATTGAT 240
Db 769 GACAAAGTGTAAATTACAGAGCTCAATCCAGATTCAGGAAACATTAAGTTGACATTGAT 828
Qy 241 TGCAGTCGAAGGTGTGTTGTTACGATCAAGGCTCCCAAGATGTTGCTCTCTCTCTCA 300
Db 829 TGCAGTCGAAGGTGTGTTGTTACGATCAAGGCTCCCAAGATGTTGCTCTCTCTCTCA 888
Qy 301 GACTGTTTCTCACTGTACTTCTGGGTAACTGGAAGAGAGCTTCACTGTTCACTG 360
Db 889 GACTGTTTCTCACTGTACTTCTGGGTAACTGGAAGAGAGCTTCACTGTTCACTG 948
Qy 361 CTGGCAGGTGGTGTGCTGAGTTCTCTGTTGTTTCCCTGCTCTGTGAAGAAATTC 420
Db 949 CTGGCAGGTGGTGTGCTGAGTTCTCTGTTGTTTCCCTGCTCTGTGAAGAAATTC 1008
Qy 421 ACTGAGTCCCTACCTGATTTCTCAGCCTTGCTTACCTGTGCAACATTTGGGCCAAC 480
Db 1009 ACTGAGTCCCTACCTGATTTCTCAGCCTTGCTTACCTGTGCAACATTTGGGCCAAC 1068
Qy 481 CGAATTTCTCCCAATCTTATCTTGCTGCGCAGCAGAGATGCTCAACAAAGAGCTGATA 540
Db 1069 CGAATTTCTCCCAATCTTATCTTGCTGCGCAGCAGAGATGCTCAACAAAGAGCTGATA 1128
Qy 541 CAGCAAGATGGATTTGTTATGTTAAATGCCAGTAACTGTGCCAAAGCTGACTTT 600
Db 1129 CAGCAAGATGGATTTGTTATGTTAAATGCCAGTAACTGTGCCAAAGCTGACTTT 1188
Qy 601 ATCCCGAGTCAATTTCCGCGTGTGCTGTAATGACGCTTTTGTGAAGAAATTTTG 660
Db 1189 ATCCCGAGTCAATTTCCGCGTGTGCTGTAATGACGCTTTTGTGAAGAAATTTTG 1248
Qy 661 CCGTGTGTGAACAATCAGTATGATTTGATGAGAAACAAAGGCTTCAATGATGTGT 720
Db 1249 CCGTGTGTGAACAATCAGTATGATTTGATGAGAAACAAAGGCTTCAATGATGTGT 1308
Qy 721 CTATGCACTGTTTACGTGGGATTCCTCGCTCCGCAACATGCTATGCTTACATCATG 780
Db 1309 CTATGCACTGTTTACGTGGGATTCCTCGCTCCGCAACATGCTTACATCATG 1368
Qy 781 AAGAGATGACATGCTTTAGATGAGAGTTTACAGATTTTGTGAAGAAAGAAACCTACT 840
Db 1369 AAGAGATGACATGCTTTAGATGAGAGTTTACAGATTTTGTGAAGAAAGAAACCTACT 1428
Qy 841 ATATCTCAAACTTCAATTTTCTGGGCAACTCTGAGACTATGAGAAAGATTAAGAAC 900
Db 1429 ATATCTCAAACTTCAATTTTCTGGGCAACTCTGAGACTATGAGAAAGATTAAGAAC 1488

Db 709 CACATTTTGAAGCCATTAAATCAACTCTCCAACTTATAGAGCAAGGTTGCAACG 768
Qy 181 GCAAGGTTAAATACAGAGCTCATCCAGCAATTACGCAAAATAGGTTGACATTGAT 240
Db 769 GCAAAAGTTTAAATACAGAGCTCATCCAGCAATTACGCAAAATAGGTTGACATTGAT 828
Qy 241 TCCAGTCAGAGGTTGATGATTACAGTCAAGAGTCCCAAGGTTGCTCTCTCTTCA 300
Db 829 TCCAGTCAGAGGTTGATGATTACAGTCAAGAGTCCCAAGGTTGCTCTCTCTTCA 888
Qy 301 GACTGTTTCTGACTGTAATCTTGAGTAACTGAGAAAGAGCTTCACTGTTCACTG 360
Db 889 GACTGTTTCTGACTGTAATCTTGAGTAACTGAGAAAGAGCTTCACTGTTCACTG 948
Qy 361 CTTCGAGGTGGTTGCTGAGTCTCTGTTGTTTCCCTGGCTCTGTGAAGAAATCC 420
Db 949 CTTCGAGGTGGTTGCTGAGTCTCTGTTGTTTCCCTGGCTCTGTGAAGAAATCC 1008
Qy 421 ACTGATGCTCCCTACCTGCAATTTCTCAGCCTTGCTTACCTGTTGCAACATTGGGCAAC 480
Db 1009 ACTGATGCTCCCTACCTGCAATTTCTCAGCCTTGCTTACCTGTTGCAACATTGGGCAAC 1068
Qy 481 CGAATCTTCCCAATCTTATCTTGAGTGGCTGCCAGAGATGCTTCAACAGAGCTGATA 540
Db 1069 CGAATCTTCCCAATCTTATCTTGAGTGGCTGCCAGAGATGCTTCAACAGAGCTGATA 1128
Qy 541 CAGAGATGAGGATGTTATGTTAAATGCGAGTATACCTGTCGCAAGCCCTGACTT 600
Db 1129 CAGAGATGAGGATGTTATGTTAAATGCGAGTATACCTGTCGCAAGCCCTGACTT 1188
Qy 601 ATCCCGAGTCTCATTTCTGCTGCTGCTGCTGAGTACAGCTTTGTGAGAAATTTTG 660
Db 1189 ATCCCGAGTCTCATTTCTGCTGCTGCTGCTGAGTACAGCTTTGTGAGAAATTTTG 1248
Qy 661 CCGTGTGTGCAATAGTAGATTGATGAAAGCAAAAGCCCTTCAATGATGTT 720
Db 1249 CCGTGTGTGCAATAGTAGATTGATGAAAGCAAAAGCCCTTCAATGATGTT 1308
Qy 721 CTAGTGCATGTTTAACTGAGATCTCCGCTCCGCCACATGCTATGCTCATCATG 780
Db 1309 CTAGTGCATGTTTAACTGAGATCTCCGCTCCGCCACATGCTATGCTCATCATG 1368
Qy 781 AAGAGATGAGCANTGCTTAAATGAAAGCTTACAGATTGTTGAAAGAAAAAGACTT 840
Db 1369 AAGAGATGAGCANTGCTTAAATGAAAGCTTACAGATTGTTGAAAGAAAAAGACTT 1428
Qy 841 ATATCTCCAACTTCAATTTCTGGGCAATCTCTGAGCTATGAGAAAGATTAAAG 900
Db 1429 ATATCTCCAACTTCAATTTCTGGGCAATCTCTGAGCTATGAGAAAGATTAAAG 1488
Qy 901 CAAACTGAGCATCAGGGCCAAAGACAAATCTCAAGCTGCTGACCTGAGAGCCAAAT 960
Db 1489 CAAACTGAGCATCAGGGCCAAAGACAAATCTCAAGCTGCTGACCTGAGAGCCAAAT 1548
Qy 961 GAACTGTCCCTGCTCTCAGAGGGTGAAGAAAGGAGACGCCCTCAGTCAACC 1020
Db 1549 GAACTGTCCCTGCTCTCAGAGGGTGAAGAAAGGAGACGCCCTCAGTCAACC 1608
Qy 1021 TGTGCCAGTCTGCTACCTCAGAGGAGAGCAAAAGCCGCTGATCCGCCAGCGTG 1080
Db 1609 TGTGCCAGTCTGCTACCTCAGAGGAGAGCAAAAGCCGCTGATCCGCCAGCGTG 1668
Qy 1081 CCCAGGTGCCAGAGGTGAGCCGCTGCTGTTAGAGACAGCCGCTGATCAGGCGCTC 1140
Db 1669 CCCAGGTGCCAGAGGTGAGCCGCTGCTGTTAGAGACAGCCGCTGATCAGGCGCTC 1728
Qy 1141 AGTGGCTGCACTTCCGAGAGAGGCTGGAAGACAGATTAAGCTTAAAGCTTCTTC 1200
Db 1729 AGTGGCTGCACTTCCGAGAGAGGCTGGAAGACAGATTAAGCTTAAAGCTTCTTC 1788
Qy 1201 TCTCTGATATCAATCAGTTTCAATTTCAAGCAAGATGAGATCTTCAATGAGCTTC 1260

Db 1789 TCTCTGATATCAATCAGTTTCAATTTCAATTCAGCCAGATGAGCATCTTACATGGCTTC 1848
Qy 1261 TCCATCAGAGATGCTTTGAAATCTACAAACCTTCCACTATCTTGAATGGACCAAC 1320
Db 1849 TCCATCAGAGATGCTTTGAAATCTACAAACCTTCCACTATCTTGAATGGACCAAC 1908
Qy 1321 AAGCTATGCCAGTTCTCCCTGTTGAGAACTATGAGAGCAACTCCGAAACAGTCT 1380
Db 1909 AAGCTATGCCAGTTCTCCCTGTTGAGAACTATGAGAGCAACTCCGAAACAGTCT 1968
Qy 1381 GATTAAGAGAGAGCCAGCATCTCCCAAGAGCTGCAAGCCGCCACTTCAAGACCCAG 1440
Db 1969 GATTAAGAGAGAGCCAGCATCTCCCAAGAGCTGCAAGCCGCCACTTCAAGACCCAG 2028
Qy 1441 AGCAAGCATTCATTCGCTGAGAACCCAGCAGAGTGGACCGCCAGAGGTCCTTTTA 1500
Db 2029 AGCAAGCATTCATTCGCTGAGAACCCAGCAGAGTGGACCGCCAGAGGTCCTTTTA 2088
Qy 1501 TCTCCACTGATGCAAGTGGAGGCTGAGAGCAATTAACAACAGCTTCTTTCCGC 1560
Db 2089 TCTCCACTGATGCAAGTGGAGGCTGAGAGCAATTAACAACAGCTTCTTTCCGC 2148
Qy 1561 CTTTCACACGAGCAGAGCACTTCAAGAGTCTGCTGCTGGGCTTTAAGGCTGGCAC 1620
Db 2149 CTTTCACACGAGCAGAGCACTTCAAGAGTCTGCTGCTGGGCTTTAAGGCTGGCAC 2208
Qy 1621 TCGATATCTTGGCCCCCAGACCTTACCCCTTCCCTGACAGCAGCTGTTATTTTGGC 1680
Db 2209 TCGATATCTTGGCCCCCAGACCTTACCCCTTCCCTGACAGCAGCTGTTATTTTGGC 2268
Qy 1681 ACAGAGTCTCACTTCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1740
Db 2269 ACAGAGTCTCACTTCTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2328
Qy 1741 GCTTACAGCTCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
Db 2329 GCTTACAGCTCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2388
Qy 1801 CAGAGGCTCAATGACAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860
Db 2389 CAGAGGCTCAATGACAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2448
Qy 1861 AAGAGTTTAAACAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
Db 2449 AAGAGTTTAAACAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2508
Qy 1921 AGGTACAGGAGAGAGCTGAGGAGAGTGGAGTCACTTCTTCTGAGGAGCATGAA 1980
Db 2509 AGGTACAGGAGAGAGCTGAGGAGAGTGGAGTCACTTCTTCTGAGGAGCATGAA 2568
Qy 1981 ATCATGAGGTTCC 1995
Db 2569 ATCATGAGGTTCC 2583

RESULT 12
ABV20978
ID ABV20978 standard; cDNA; 5145 BP.
XX
XX ABV20978;
DT 13-SEP-2002 (first entry)
XX
XX Human prostate expression marker cDNA 20969.
DE
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW Pharmacogenomic marker; gene; ss.
XX
XX Homo sapiens.
OS
XX
PN W0200160860-A2.
XX
PD 23-AUG-2001.

[illegible]

QY	481	GGAAATTCCTCCCAATCTTTATCTTGGCGCCAGGAGATGTCCTCAACAAGACCTGATA	540
Db	1069	CGAAATCTTCCCAATCTTTATCTTGGCGCCAGGAGATGTCCTCAACAAGACCTGATG	1128
QY	541	CAGACGAATGGGATTTGGTTATGTATGTAAATCCACAGCTATACCTGTCCAAAGCCTGACTTT	600
Db	1129	CAGCGAATGGGATTTGGTTATGTGTAAATGCCAGCAATACCTGTCCAAAGCCTGACTTT	1188
QY	601	ATCCCGGAGTCAATTTCTCGCGTGCCTGTGAAATGACAGCTTTTGTGAAAAATTTTG	660
Db	1189	ATCCCGGAGTCAATTTCTCGCGTGCCTGTGAAATGACAGCTTTTGTGAAAAATTTTG	1248
QY	661	CCGTGTGTGGCAAAATCAGTATGATTTTCATTTGAGAAAGCAAAAGCCTCCATATGATGTCTT	720
Db	1249	CCGTGTGTGGCAAAATCAGTATGATTTTCATTTGAGAAAGCAAAAGCCTCCAAATGATGTGT	1308
QY	721	CTAGTGCATCTATTAGCTGGGAGATCTCCCGCGCTCCGSCACATCGCTATCGCTTACATCATG	780
Db	1309	CTAGTGCATCTATTAGCTGGGATCTCCCGCGCTCCGSCACATCGCTATCGCTTACATCATG	1368
QY	781	AAGAGATGACATGTCTTTAGATGAAGCTTACAGATTTTGTGAAAAAGAAAAAGACTACT	840
Db	1369	AAGAGATGACATGTCTTTAGATGAAGCTTACAGATTTTGTGAAAAAGAAAAAGACTACT	1428
QY	841	ATATCTCCAACTTCAATTTTCTGGGCGCACTCTCGACATTTGAGAAAGATTTAAGAC	900
Db	1429	ATATCTCCAACTTCAATTTTCTGGGCGCACTCTCGACATTTGAGAAAGATTTAAGAC	1488
QY	901	CAGACTGGAGCATAGGGCCCAAAGCAAACTCAAGCGCTGCACCTGGAGAACCCAAT	960
Db	1489	CAGACTGGAGCATAGGGCCCAAAGCAAACTCAAGCTGCACCTGGAGAACCCAAT	1548
QY	961	GAACTGTCCCTGCTGTCTCAGAGGGTGTGACAGAAAAAGCAGACGCCCTCAGTCCACC	1020
Db	1549	GAACTGTCCCTGCTGTCTCAGAGGGTGTGACAGAAAAAGCAGACGCCCTCAGTCCACC	1608
QY	1021	TGTGCGCATCTTGCTACCTCAGAGGCGACAGAGCAAAAGGCCCTGTGCATTCCTCCAGCGTG	1080
Db	1609	TGTGCGCATCTTGCTACCTCAGAGGCGACAGAGCAAAAGGCCCTGTGCATTCCTCCAGCGTG	1668
QY	1081	CCGACGCGGCCCAAGGTGACGCGGTGCTTTAAGAGCAAGCCGCTGGTTACAGGCGCTC	1140
Db	1669	CCGACGCGGCCCAAGGTGACGCGGTGCTTTAAGAGCAAGCCGCTGGTTACAGGCGCTC	1728
QY	1141	AGTGGGCTGCACCTGTCCGCGACAGAGCTGGAAAGACAGCATTAAGCTCAAGCTTTCTCTC	1200
Db	1729	AGTGGGCTGCACCTGTCCGCGACAGAGCTGGAAAGACAGCATTAAGCTCAAGCTTTCTCTC	1788
QY	1201	TCTCTGATATCAAAATCAGTTTCATTTACGCCAGATGGCAGATCTTTACATGCGTTC	1260
Db	1789	TCTCTGATATCAAAATCAGTTTCATTTACGCCAGATGGCAGATCTTTACATGCGTTC	1848
QY	1261	TCTCTATCAGAAAGTGCCTTTGGAATATCTAACAACTTCCACTACTCTGTGATGGGACCAAC	1320
Db	1849	TCTCTATCAGAAAGTGCCTTTGGAATATCTAACAACTTCCACTACTCTGTGATGGGACCAAC	1908
QY	1321	AAGCTATGCCAGTTCTCCCTGTTCAGAGAACTATCGGAGGAGACTCCCGAAACCAAGTCTCT	1380
Db	1909	AAGCTATGCCAGTTCTCCCTGTTCAGAGAACTATCGGAGGAGACTCCCGAAACCAAGTCTCT	1968
QY	1381	GATTAAGAGAGAGCCAGCATCCCGAAGAGCTGACAGCCGCGAGGCGCTTCAGACAGCCAG	1440
Db	1969	GATTAAGAGAGAGCCAGCATCCCGAAGAGCTGACAGCCGCGAGGCGCTTCAGACAGCCAG	2028
QY	1441	AGCAAGGATTTGGATTGGGTCCAGAACCAAGACAGCATGGGACCCGCGCAGAGGTCCCTTTTA	1500
Db	2029	AGCAAGGATTTGGATTGGGTCCAGAACCAAGACAGCATGGGACCCGCGCAGAGGTCCCTTTTA	2088
QY	1501	TCTCCACTGCATCAAAATGGGAGCGTGGAGGACAAATTACACACCAAGCTTCTTTTTCGGC	1560
Db	2089	TCTCCACTGCATCAAAATGGGAGCGTGGAGGACAAATTACACACCAAGCTTCTTTTTCGGC	2148
QY	1561	CTTTTCCACAGCCGACACCTTCAGAGAGTCTGTGCGCTTGAAGGCTGTGCAC	1620


```

Db      2149 CTTTCCACGACGACGACGACCTCAGAGAGTGTGAGCTTGAGCCTTAAAGGCTGAC 2208
QY      1621 TCGGATATCTTGAGCCGCCGAGACCTTACCTTCCCTGACGACAGCTGTAATTTTGC 1680
Db      2209 TCGGATATCTTGAGCCGCCGAGACCTTACCTTCCCTGACGACAGCTGTAATTTTGC 2268
QY      1681 ACGAGTCTCAGACTTCTACTTGTGCTCAGGCACTTACGAGAGGAGTGCAGTTACT 1740
Db      2269 ACGAGTCTCAGACTTCTACTTGTGCTCAGGCACTTACGAGAGGAGTGCAGTTACT 2328
QY      1741 GCGTACAGCTGACGACGAGCTGCGCAGTGGCGAGACCAAGTATTTCTGCGCAGGCG 1800
Db      2329 GCGTACAGCTGACGACGAGCTGCGCAGTGGCGAGACCAAGTATTTCTGCGCAGGCG 2388
QY      1801 CAGAGCGCAAGTACAGAGCTGACTGCGCGGAGCTGGCATGAGAGAGCCCTTTGAA 1860
Db      2389 CAGAGCGCAAGTACAGAGCTGACTGCGCGGAGCTGGCATGAGAGAGCCCTTTGAA 2448
QY      1861 AAGCAGTTTAAACGACGAGCTGCCAATGGAATTTGAGAGAGCATCTGTCAAGAAC 1920
Db      2449 AAGCAGTTTAAACGACGAGCTGCCAATGGAATTTGAGAGAGCATCTGTCAAGAAC 2508
QY      1921 AGGTCACGGAAGAGTGGGGAAGTGGGAGTCAAGTCTTTCGGGACAGATGAA 1980
Db      2509 AGGTCACGGAAGAGTGGGGAAGTGGGAGTCAAGTCTTTCGGGACAGATGAA 2568
QY      1981 ATCATTGAGGTCTCC 1995
Db      2569 ATCATTGAGGTCTCC 2583

```

RESULT 13

ABV21092
ID ABV21092 standard; cDNA; 5145 BP.

XX AC ABV21092;

XX DT 13-SEP-2002 (first entry)

XX DE Human prostate expression marker cDNA 21083.

XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.

XX OS Homo sapiens.

XX PN W0200160860-A2.

XX PD 23-AUG-2001.

XX PP 20-FEB-2001; 2001WO-US005171.

XX PR 17-FEB-2000; 2000US-0183319P.

XX PR 16-MAR-2000; 2000US-0189862P.

XX PR 25-MAY-2000; 2000US-0207454P.

XX PR 09-JUN-2000; 2000US-0211314P.

XX PR 18-JUL-2000; 2000US-0219077P.

XX PR 13-DEC-2000; 2000US-0255281P.

XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX PI Schlegel R, Endege WO, Monahan JE;

XX DR WPI; 2001-662795/76.

XX PT Novel isolated nucleic acid molecule associated with cancerous state of
XX prostate cells and correlating with presence of prostate cancer, useful
XX for detecting presence of prostate cancer, stage of prostate cancer.
XX Claim 1; Page 3485; 11750pp; English.
XX The invention relates to an isolated nucleic acid molecule (I) comprising

CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or incidence of prostate cancer in a patient
CC ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker

Sequence 5145 BP; 1408 A; 1135 C; 1253 G; 1346 T; 0 U; 3 Other;

Query Match 94.9%; Score 1893; DB 5; Length 5145;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

QY      1 ATGGCCCATGAGATGATGGAATCAATTTGTTACTGAGAGTGGTGGCTTCTGGAA 60
Db      589 ATGGCCCATGAGATGATGGAATCAATTTGTTACTGAGAGTGGTGGCTTCTGGAA 648
QY      61 AGTGGACGGAAGAAAGTCTGTAATGATGACCGGCGCATTTTGGAAATACATATCCTC 120
Db      649 AGTGGACGGAAGAAAGTCTGTAATGATGACCGGCGCATTTTGGAAATACATATCCTC 708
QY      121 CACATTTTGAAGCCATTATATCAACTGCTCCAGCTTATGAGAGCGAAGTTGCAACAG 180
Db      709 CACATTTTGAAGCCATTATATCAACTGCTCCAGCTTATGAGAGCGAAGTTTGCACAG 768
QY      769 GACAAAGTGTATATATACAGAGCTCATCCAGATTACAGGAACAATAGTTGACATTGAT 828
Db      829 TGCAGTCAAGAGTGTATATATACAGATCAAGCTCCAAAGATGTGCTCTCTCTTCA 888
QY      241 TGCAGTCAAGAGTGTATATACAGATCAAGCTCCAAAGATGTGCTCTCTCTTCA 300
Db      889 GACGTGTTTCTCAGCTGTAATTTGATGATGATGATGATGATGATGATGATGATGAT 948
QY      301 GACGTGTTTCTCAGCTGTAATTTGATGATGATGATGATGATGATGATGATGATGAT 360
Db      949 CTTGCAGGAGGATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1008
QY      421 ACTCTAGTCCCTACCTGATTTCTCAGGCTGTTACCTGTTGCCAATTTGGGCAACC 480
Db      1009 ACTCTAGTCCCTACCTGATTTCTCAGGCTGTTACCTGTTGCCAATTTGGGCAACC 1068
QY      481 CGAATTTCTCCCAATCTTATCTTGGCTGCCAGAGATGCTTCAACAAGAGCTGATA 540
Db      1069 CGAATTTCTCCCAATCTTATCTTGGCTGCCAGAGATGCTTCAACAAGAGCTGATA 1128
QY      541 CAGCAGAAATGGATGTTATGTTAAATGTCAGCTATACCTGTCCAAAGCTTGAATTT 600
Db      1129 CAGCAGAAATGGATGTTATGTTAAATGTCAGCTATACCTGTCCAAAGCTTGAATTT 1188
QY      601 ATCCCGAGTCTCATTTCTGCTGCTGCTGGAATGACAGCTTTGTGAGAAATTTTG 660
Db      1189 ATCCCGAGTCTCATTTCTGCTGCTGCTGGAATGACAGCTTTGTGAGAAATTTTG 1248
QY      661 CCGTGGTTGGAACAATCAGTATGATTTCAATTGAGAAAGCAAAAGCTTCAATGATGT 720
Db      1249 CCGTGGTTGGAACAATCAGTATGATTTCAATTGAGAAAGCAAAAGCTTCAATGATGT 1308
QY      721 CTATGCACTGTTTATGCTGGATCTCCGCTCCGACATGCTATGCTTACATCATG 780
Db      1309 CTATGCACTGTTTATGCTGGATCTCCGCTCCGACATGCTATGCTTACATCATG 1368
QY      781 AAGAGATGGAATGCTTATGATGAGCTTCAAGATTTGTGAGAAAGAAAGAAAGCTTACT 840
Db      1369 AAGAGATGGAATGCTTATGATGAGCTTCAAGATTTGTGAGAAAGAAAGAAAGCTTACT 1428

```

```
QY 841 ATATCTCAAACTTCAATTTTCTGGGCAACTCTGACTATGAGAAAGATTTAAGAAC 900
    |||||
Db 1429 ATATCTCAAACTTCAATTTTCTGGGCAACTCTGACTATGAGAAAGATTTAAGAAC 1488
QY 901 CAGACTGGAGCATCAGGGCCAAAGAGCAAACTCAGCTGTGCACTCGAGAAAGCCAAAT 960
    |||||
Db 1489 CAGACTGGAGCATCAGGGCCAAAGAGCAAACTCAGCTGTGCACTCGAGAAAGCCAAAT 1548
QY 961 GAAACTGTCTCTGCTGTCTCAAGAGGTGACAGAGAAAAGCAGACGCCCTCACTCCACC 1020
    |||||
Db 1549 GAAACTGTCTCTGCTGTCTCAAGAGGTGACAGAGAAAAGCAGACGCCCTCACTCCACC 1608
QY 1021 TGTGGCGACTCTGTACTCTCAGAGGCGAGCAAAAGGCCGTGCACTCCGCGCAGGTG 1080
    |||||
Db 1609 TGTGGCGACTCTGTACTCTCAGAGGCGAGCAAAAGGCCGTGCACTCCGCGCAGGTG 1668
QY 1081 CCCAGCGTGGCCAGCGCTGACAGCGCTGCTGTAGAGAGACGCCGCTGTACAGAGGCTC 1140
    |||||
Db 1669 CCCAGCGTGGCCAGCGCTGACAGCGCTGCTGTAGAGAGACGCCGCTGTACAGAGGCTC 1728
QY 1141 AGTGGGCTGCACCTGTCTCCGACAGAGGCTGGAAGACAGCAATTAAGTCAAGCGTCTTC 1200
    |||||
Db 1729 AGTGGGCTGCACCTGTCTCCGACAGAGGCTGGAAGACAGCAATTAAGTCAAGCGTCTTC 1788
QY 1201 TCTCTGGATATCAATGATTTTCAATATATTCAGCCAGCATGGCAGCATCTTACATAGGCTTC 1260
    |||||
Db 1789 TCTCTGGATATCAATGATTTTCAATATATTCAGCCAGCATGGCAGCATCTTACATAGGCTTC 1848
QY 1261 TCCTCATCAGAGATGCTTTGGAATATCTACAAACCTTTCACACTCTGTGATGGAGCAAC 1320
    |||||
Db 1849 TCCTCATCAGAGATGCTTTGGAATATCTACAAACCTTTCACACTCTGTGATGGAGCAAC 1908
QY 1321 AAGCTATGCCAGTTCTTCCCTGTTCAAGAACTATCGSAGCAGACTCCGAAACCAAGTCTT 1380
    |||||
Db 1909 AAGCTATGCCAGTTCTTCCCTGTTCAAGAACTATCGSAGCAGACTCCGAAACCAAGTCTT 1968
QY 1381 GATTAAGGAGAGAGCAGCATCCCAAGAAAGCTGACAGCGCCGCTTACAGACGCGAG 1440
    |||||
Db 1969 GATTAAGGAGAGAGCAGCATCCCAAGAAAGCTGACAGCGCCGCTTACAGACGCGAG 2028
QY 1441 AGCAAGCGATTTGATTCGTCTAGAACAGCAGCAGAGTGGACCCGCGAGAGGTCTCTTTTA 1500
    |||||
Db 2029 AGCAAGCGATTTGATTCGTCTAGAACAGCAGCAGAGTGGACCCGCGAGAGGTCTCTTTTA 2088
QY 1501 TCTTCACTGCATCGAAGTGGAGCGTGGAGACAATTACCAACACAGCTTCTTTTCGGC 1560
    |||||
Db 2089 TCTTCACTGCATCGAAGTGGAGCGTGGAGACAATTACCAACACAGCTTCTTTTCGGC 2148
QY 1561 CTTTTCACCAAGCAGAGCAGCACTTCAAGAAATCTGCTGGCTTGAAGGCTGGCAC 1620
    |||||
Db 2149 CTTTTCACCAAGCAGAGCAGCACTTCAAGAAATCTGCTGGCTTGAAGGCTGGCAC 2208
QY 1621 TCGGATATCTTTGGCCCCCAGACCTTACCCCTTCCCTGACAGCAGCTGTAATTTTGGC 1680
    |||||
Db 2209 TCGGATATCTTTGGCCCCCAGACCTTACCCCTTCCCTGACAGCAGCTGTAATTTTGGC 2268
QY 1681 ACAGAGTCTTCAACATTTACTGTGCTTACGCCATCTTACGAGAGCAGTCCAGTTACTCT 1740
    |||||
Db 2269 ACAGAGTCTTCAACATTTACTGTGCTTACGCCATCTTACGAGAGCAGTCCAGTTACTCT 2328
QY 1741 GCTTACAGCTGACAGCAGCTGCCCACTTGGGAGACCAAGTCTAATTTCTGTGGCAGGCGG 1800
    |||||
Db 2329 GCTTACAGCTGACAGCAGCTGCCCACTTGGGAGACCAAGTCTAATTTCTGTGGCAGGCGG 2388
QY 1801 CAGAAAGCAAGTGAAGAGCTGACTCGCGCGAGAGTGCATGAAGAGAGCCCTTTTGA 1860
    |||||
Db 2389 CAGAAAGCAAGTGAAGAGCTGACTCGCGCGAGAGCTGCATGAAGAGAGCCCTTTTGA 2448
QY 1861 AAGCAGTTTAAAGCAGAGCTGCCTCAATGAAATTTTGAAGAGCATGTTCAGAGAAC 1920
    |||||
Db 2449 AAGCAGTTTAAAGCAGAGCTGCCTCAATGAAATTTTGAAGAGCATGTTCAGAGAAC 2508
```

```
QY 1921 AGGTCAAGGAGAGAGCTGGGAAAGTGGCAGTCACTAGCTTTTGGGACAGATGAA 1980
    |||||
Db 2509 AGGTCAAGGAGAGAGCTGGGAAAGTGGCAGTCACTAGCTTTTGGGACAGATGAA 2568
QY 1981 ATCATTTGAGGTCTCC 1995
    |||||
Db 2569 ATCATTTGAGGTCTCC 2583
    |||||

RESULT 14
ABV21312
ID ABV21312 standard; cDNA; 5145 BP.
XX
AC ABV21312;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 21303.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
    pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN MO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-018319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
DX Novel isolated nucleic acid molecule associated with cancerous state of
    prostate cells and correlating with presence of prostate cancer, useful
    for detecting presence of prostate cancer, stage of prostate cancer.
XX
PT Claim 1; Page 3539; 11750pp; English.
XX
PS
XX
CC The invention relates to an isolated nucleic acid molecule (1) comprising
    a nucleotide sequence given in Tables 1-9 (ABV0010-ABV62213) of the
    specification or its complement. (1) is useful for: (a) assessing whether
    a patient is afflicted with prostate cancer; (b) monitoring the
    CC progression of prostate cancer in a patient; (c) assessing the efficacy
    of a test compound to inhibit prostate cancer in a patient; (d) assessing
    the efficacy of a therapy for inhibiting prostate cancer in a patient;
    CC (e) selecting a composition for inhibiting prostate cancer in a patient;
    CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
    CC determining whether prostate cancer has metastasized in a patient; (h)
    CC assessing the aggressiveness or indolence of prostate cancer in a patient
    CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
    XX
SQ Sequence 5145 BP; 1408 A; 1135 C; 1253 G; 1346 T; 0 U; 3 Other;
    Query Match 94.9%; Score 1893; DB 5; Length 5145;
    Best Local Similarity 99.9%; Pred. No. 0;
    Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCCCATGAGATGATGGAATCTCAATTTGTTACTGAGAGGTGTGGCTCTGCTGAA 60
    |||||
Db 589 ATGGCCCATGAGATGATGGAATCTCAATTTGTTACTGAGAGGTGTGGCTCTGCTGAA 648
    |||||
QY 61 AGTGAACGGAAAAAGTGTCTGCTAATTGATTAACCGGCCATTTGTGGAATACATCATCC 120
    |||||
```

|||||
Db 649 AGTGAACGGAAAAATGGCTGAATTTAGAGCCGGCCATTTGGAAATACAAATACATCC 708
|||
Oy 121 CACATTTTGGAAACCTTAATATCACTGCTCCAGCTTATGAGGAAAGTTGGAACG 180
|||
Db 709 CACATTTTGGAAACCTTAATATCACTGCTCCAGCTTATGAGGAAAGTTGGAACG 768
|||
Oy 181 GACAAAGTGTAAATACAGAGCTCATCCAGCATTCAGCAAACTAATAGTTGACATTAAT 240
|||
Db 769 GACAAAGTGTAAATACAGAGCTCATCCAGCATTCAGCAAACTAATAGTTGACATTAAT 828
|||
Oy 241 TGCAGTCAGAAAGTTGAGTTAGATCAATCAAAAGTCCCAAGATGTTGCTCTCTTCA 300
|||
Db 829 TGCAGTCAGAAAGTTGAGTTAGATCAATCAAAAGTCCCAAGATGTTGCTCTCTTCA 888
|||
Oy 301 GACTGTTTCTCACTGACTTCTGGGTTAACTGAGAGAGCTTCAACTCTGTTCACTG 360
|||
Db 889 GACTGTTTCTCACTGACTTCTGGGTTAACTGAGAGAGCTTCAACTCTGTTCACTG 948
|||
Oy 361 CTTCAGAGTGGGTTTGTGCTGAGATTCTCTGTTTCTCTGCTCTGTTGAGAGAAATCC 420
|||
Db 949 CTTCAGAGTGGGTTTGTGCTGAGATTCTCTGTTTCTCTGCTCTGTTGAGAGAAATCC 1008
|||
Oy 421 ACTCTAGTCCCTTACCTGCAATTTCTGACCTTGTCTTACCTGTTGAGAGAGAGCT 480
|||
Db 1009 ACTCTAGTCCCTTACCTGCAATTTCTGACCTTGTCTTACCTGTTGAGAGAGAGCT 1068
|||
Oy 481 GCAATTTCTCCCAATCTTTATCTTGGCTGCACAGAGATGTCCTCAACAAGAGCTGATA 540
|||
Db 1069 GCAATTTCTCCCAATCTTTATCTTGGCTGCACAGAGATGTCCTCAACAAGAGCTGATA 1128
|||
Oy 541 CAGCAGATGGGATTTGTTATGTTTAATGCCAGCTATACCTGTCCAAAGCTGACTTT 600
|||
Db 1129 CAGCAGATGGGATTTGTTATGTTTAATGCCAGCTATACCTGTCCAAAGCTGACTTT 1188
|||
Oy 601 ATCCCGCAGTCTCATTTCTGCGCGTGCCTGCTGGAATGACAGCTTTTGTGAGAAATTTTG 660
|||
Db 1189 ATCCCGCAGTCTCATTTCTGCGCGTGCCTGCTGGAATGACAGCTTTTGTGAGAAATTTTG 1248
|||
Oy 661 CCGTGTGTGACAAATCAGTAGATTTCATTGAGAAAGCAAAAGCTCCAAATGATGTTT 720
|||
Db 1249 CCGTGTGTGACAAATCAGTAGATTTCATTGAGAAAGCAAAAGCTCCAAATGATGTTT 1308
|||
Oy 721 CTAAGTCACCTGTTAGCTGGGATCTCCGCTCCGSCCACTCGTATGCTCATCATGAG 780
|||
Db 1309 CTAAGTCACCTGTTAGCTGGGATCTCCGCTCCGSCCACTCGTATGCTCATCATGAG 1368
|||
Oy 781 AAGAGATGACATGTCCTTAGATGAAGCTTACAGATTGTAAGAAAAAAGAAAGCTACT 840
|||
Db 1369 AAGAGATGACATGTCCTTAGATGAAGCTTACAGATTGTAAGAAAAAAGAAAGCTACT 1428
|||
Oy 841 ATATCTCCAAATCTCAATTTTCTGGGCCCACTCTCGACATATGAGAAAGATTAAAGAC 900
|||
Db 1429 ATATCTCCAAATCTCAATTTTCTGGGCCCACTCTCGACATATGAGAAAGATTAAAGAC 1488
|||
Oy 901 CAGACTGAGAGATCAGGGGCCAAAGAGCAAACTCAAGCTGCTGACCTGAGAAAGCAAAAT 960
|||
Db 1489 CAGACTGAGAGATCAGGGGCCAAAGAGCAAACTCAAGCTGCTGACCTGAGAAAGCAAAAT 1548
|||
Oy 961 GAACTGTCTCCTGTCTCAGAGGGTGAACAGAAAGAGAGAGCCCTTCACTGCAACC 1020
|||
Db 1549 GAACTGTCTCCTGTCTCAGAGGGTGAACAGAAAGAGAGAGCCCTTCACTGCAACC 1608
|||
Oy 1021 TGTGCGCACTGTCTCAGAGGACAGCAAGAAAGCCCGTGCATCCCGCAGAGGTG 1080
|||
Db 1609 TGTGCGCACTGTCTCAGAGGACAGCAAGAAAGCCCGTGCATCCCGCAGAGGTG 1668
|||
Oy 1081 CCCAGCTGCCCCAGGCTGAGCCGCTGCTGTTAGAGGACAGCCCGCTGTTACAGCGCTC 1140
|||
Db 1669 CCCAGCTGCCCCAGGCTGAGCCGCTGCTGTTAGAGGACAGCCCGCTGTTACAGCGCTC 1728
|||
Oy 1141 AGTGGGCTGACCTGTCTCGCAGACAGGCTGGAAGACAGCAATTAAGCTTAAGCTTCTCTC 1200
|||

Db 1729 AGTGGCTGCACTGTCCGACAGCAGGCTGGAAGACAGCAATTAAGCTCAAGGCTTCTTC 1788
|||
Oy 1201 TCTCTGATATCAAAATCATGTTTCAATATTCAGCCAGATGCGACATCTTACATGGCTTC 1260
|||
Db 1789 TCTCTGATATCAAAATCATGTTTCAATATTCAGCCAGATGCGACATCTTACATGGCTTC 1848
|||
Oy 1261 TCTCTATCAAAAGATGTTTGGAAATGATCAAACTTCCACTCTGATGAGGACCAAC 1320
|||
Db 1849 TCTCTATCAAAAGATGTTTGGAAATGATCAAACTTCCACTCTGATGAGGACCAAC 1908
|||
Oy 1321 AAGCTATGCAAGTTCTCCCTGTTTCAAGAACTATGGAAGACATCTCCGAAACCACTCT 1380
|||
Db 1909 AAGCTATGCAAGTTCTCCCTGTTTCAAGAACTATGGAAGACATCTCCGAAACCACTCT 1968
|||
Oy 1381 GATTAAGAGGAAGCAGACATCTCCCAAGAAAGCTGACAGCCGAGGCTTACAGAGCAG 1440
|||
Db 1969 GATTAAGAGGAAGCAGACATCTCCCAAGAAAGCTGACAGCCGAGGCTTACAGAGCAG 2028
|||
Oy 1441 AGCAAGGATTTGCAATGCGGTCAAGAACAGCAGAGAGTGGCACCGCCAGAGGTCCTTTTA 1500
|||
Db 2029 AGCAAGGATTTGCAATGCGGTCAAGAACAGCAGAGAGTGGCACCGCCAGAGGTCCTTTTA 2088
|||
Oy 1501 TCTCCACTGATGCAAGTGGAGCTGAGAGACATTAACACACAGCTTCTTTTCGCG 1560
|||
Db 2089 TCTCCACTGATGCAAGTGGAGCTGAGAGACATTAACACACAGCTTCTTTTCGCG 2148
|||
Oy 1561 CTTTCCACACAGCAGCAGACCTTCAGAGAGTCTGCTGAGGCTTAAAGGCTGGCAC 1620
|||
Db 2149 CTTTCCACACAGCAGCAGACCTTCAGAGAGTCTGCTGAGGCTTAAAGGCTGGCAC 2208
|||
Oy 1621 TCGATATCTTGGGCCCCCAGACCTTAACCCCTTCCCTGACACAGAGCTGGATTTTGGC 1680
|||
Db 2209 TCGATATCTTGGGCCCCCAGACCTTAACCCCTTCCCTGACACAGAGCTGGATTTTGGC 2268
|||
Oy 1681 ACAGAGTCTTCACTATCTTCTGCTCAGCCATCTAAGGAGCAGTCAAGTTACTCT 1740
|||
Db 2269 ACAGAGTCTTCACTATCTTCTGCTCAGCCATCTAAGGAGCAGTCAAGTTACTCT 2328
|||
Oy 1741 GCTTACAGCTGACGACGCTGCCCATTTGCGGAGACCAAGTCTATTCTGTGCGAGCGG 1800
|||
Db 2329 GCTTACAGCTGACGACGCTGCCCATTTGCGGAGACCAAGTCTATTCTGTGCGAGCGG 2388
|||
Oy 1801 CAGAAAGCAAGTACAGAGCTGACTCTGCGGCGGAGCTGGCATTAAGAGAGCCCTTTGAA 1860
|||
Db 2389 CAGAAAGCAAGTACAGAGCTGACTCTGCGGCGGAGCTGGCATTAAGAGAGAGCCCTTTGAA 2448
|||
Oy 1861 AAGCATTTTAAACGAGAAAGCTGCCAAATGGAATTTGAGAGACATCATGTCAAGAAC 1920
|||
Db 2449 AAGCATTTTAAACGAGAAAGCTGCCAAATGGAATTTGAGAGACATCATGTCAAGAAC 2508
|||
Oy 1921 AGGTACCGGAAAGAGCTGGGAAAGTGGGAGTCACTAGCTTTTGGGACGATGGAA 1980
|||
Db 2509 AGGTACCGGAAAGAGCTGGGAAAGTGGGAGTCACTAGCTTTTGGGACGATGGAA 2568
|||
Oy 1981 ATCATTTAGGTTCC 1995
|||
Db 2569 ATCATTTAGGTTCC 2583
|||
RESULT 15
ABV21316
ID ABV21316 standard; cDNA; 5145 BP.
XX
XX ABV21316;
XX
DT 13-SBP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 21307.
XX
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KM pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.

XX WO200160860-A2.
 PN 23-AUG-2001.
 PD 20-FEB-2001; 2001WO-US005171.
 XX 17-FEB-2000; 2000US-018319P.
 PR 16-MAR-2000; 2000US-0189862P.
 PR 25-MAY-2000; 2000US-0207454P.
 PR 09-JUN-2000; 2000US-0211314P.
 PR 18-JUL-2000; 2000US-0219007P.
 PR 13-DEC-2000; 2000US-0255281P.
 XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 PA Schlegel R, Endege WO, Monahan JR;
 PI WPI; 2001-662795/76.
 DR Novel isolated nucleic acid molecule associated with cancerous state of
 PT prostate cells and correlating with presence of prostate cancer, useful
 PT for detecting presence of prostate cancer, stage of prostate cancer.
 XX
 XX Claim 1, Page 3540; 11750pp; English.
 PS The invention relates to an isolated nucleic acid molecule (I) comprising
 CC a nucleotide sequence given in Tables 1-9 (ABV0010-ABV62213) of the
 CC specification or its complement. (I) is useful for: (a) assessing whether
 CC a patient is afflicted with prostate cancer; (b) monitoring the
 CC progression of prostate cancer in a patient; (c) assessing the efficacy
 CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
 CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
 CC (e) selecting a composition for inhibiting prostate cancer in a patient;
 CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
 CC determining whether prostate cancer has metastasized in a patient; (h)
 CC assessing the aggressiveness or indolence of prostate cancer in a patient
 CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
 CC
 XX Sequence 5145 BP; 1408 A; 1135 C; 1253 G; 1346 T; 0 U; 3 Other;
 SQ
 Query Match 94.9%; Score 1893; DB 5; Length 5145;
 Best local Similarity 99.9%; Pred. No. 0;
 Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 421 ACTGATGCTTACCTGATCTTCTCAGCCTTGTCTTCTGTTGCCAATGCGGCAACC 480
 |||||
 DB 1009 ACTGATGCTTACCTGATCTTCTCAGCCTTGTCTTCTGTTGCCAATGCGGCAACC 1068
 |||||
 QY 481 CGAATCTTCCCAATCTTTATCTTGGCTGCCAGGAGATGCTCTCAACAGAGCTGAT 540
 |||||
 DB 1069 CGAATCTTCCCAATCTTTATCTTGGCTGCCAGGAGATGCTCTCAACAGAGCTGAT 1128
 |||||
 QY 541 CAGCAAAATGGATGTTATGTTTAAATGCAAGTATACCTGTCCAAAGCCTGACTT 600
 |||||
 DB 1129 CAGCAAAATGGATGTTATGTTTAAATGCAAGTATACCTGTCCAAAGCCTGACTT 1188
 |||||
 QY 601 ATCCCGAGTCTCATTTCTGCGTGGCTGTGAATGACAGCTTTGTGAGAAATTTTG 660
 |||||
 DB 1189 ATCCCGAGTCTCATTTCTGCGTGGCTGTGAATGACAGCTTTGTGAGAAATTTTG 1248
 |||||
 QY 661 CCGTGGTGGACAAATCACTAGATTTTCAATTGAGAAAGCAAAAGCCTCCAAATGATGTT 720
 |||||
 DB 1249 CCGTGGTGGACAAATCACTAGATTTTCAATTGAGAAAGCAAAAGCCTCCAAATGATGTT 1308
 |||||
 QY 721 CTATGTCATCTTTTACCTGGATCTCCGCTCCGCCCACTATGCTATCGCTTACATG 780
 |||||
 DB 1309 CTATGTCATCTTTTACCTGGATCTCCGCTCCGCCCACTATGCTATCGCTTACATG 1368
 |||||
 QY 781 AAGAGATGACATGCTTTAGATGAAGCTTACAGATTTGTGAAAAGAAAAGACCTACT 840
 |||||
 DB 1369 AAGAGATGACATGCTTTAGATGAAGCTTACAGATTTGTGAAAAGAAAAGACCTACT 1428
 |||||
 QY 841 ATATCTCCAACTTCAATTTTCTGGGCAACTCTTGGACTATGAGAAAGATTAAGAC 900
 |||||
 DB 1429 ATATCTCCAACTTCAATTTTCTGGGCAACTCTTGGACTATGAGAAAGATTAAGAC 1488
 |||||
 QY 901 CAGCTGAGCATCAGGCGCAAGCAAACTCAAGCTGCTGACCTGAGAGAGCCAAAT 960
 |||||
 DB 1489 CAGCTGAGCATCAGGCGCAAGCAAACTCAAGCTGCTGACCTGAGAGAGCCAAAT 1548
 |||||
 QY 961 GAACCTGCTCCCTGCTCTCAGAGGGTGAAGCAAAAGGAGAGCGCCCTCAGTCCACC 1020
 |||||
 DB 1549 GAACCTGCTCCCTGCTCTCAGAGGGTGAAGCAAAAGGAGAGCGCCCTCAGTCCACC 1608
 |||||
 QY 1021 TGTGCCAGCTTGTCTACCTCAGAGGAGAGCAAAAGGCGGTGATCCCGCAGCGTG 1080
 |||||
 DB 1609 TGTGCCAGCTTGTCTACCTCAGAGGAGAGCAAAAGGCGGTGATCCCGCAGCGTG 1668
 |||||
 QY 1081 CCCAGGCTCCAGCGCTGACCGCTGCTTTAGAGAGAGCGCGCTGTAACAGGCGCTC 1140
 |||||
 DB 1669 CCCAGGCTCCAGCGCTGACCGCTGCTTTAGAGAGAGCGCGCTGTAACAGGCGCTC 1728
 |||||
 QY 1141 AGTGGGCTGCACTGTCCGAGAGAGGCTGGAAGACAGCAATTAAGCTCAAGGCTCTTC 1200
 |||||
 DB 1729 AGTGGGCTGCACTGTCCGAGAGAGGCTGGAAGACAGCAATTAAGCTCAAGGCTCTTC 1788
 |||||
 QY 1201 TCTCTGATATCAATCACTTTCTATATTCAGCCAGCATGAGCATCTTACATGCTTC 1260
 |||||
 DB 1789 TCTCTGATATCAATCACTTTCTATATTCAGCCAGCATGAGCATCTTACATGCTTC 1848
 |||||
 QY 1261 TCTCTATCAGAAAGATCTTTGAAATCTACAAACCTTCCACTACTCTGATGGAGCAAC 1320
 |||||
 DB 1849 TCTCTATCAGAAAGATCTTTGAAATCTACAAACCTTCCACTACTCTGATGGAGCAAC 1908
 |||||
 QY 1321 AAGCTATGCAAGTTCCTCCCTGTTCAGGAAGTATGGAGAGAGCTCCGAAACAGTCCCT 1380
 |||||
 DB 1909 AAGCTATGCAAGTTCCTCCCTGTTCAGGAAGTATGGAGAGAGCTCCGAAACAGTCCCT 1968
 |||||
 QY 1381 GATTAAGAGAAAGCAGCATCTCCCAAGAGCTGACAGCCGAGGCTTACAGACAGCAG 1440
 |||||
 DB 1969 GATTAAGAGAAAGCAGCATCTCCCAAGAGCTGACAGCCGAGGCTTACAGACAGCAG 2028
 |||||
 QY 1441 AGCAACGATTCATTCGATTCGATCAGAACACAGACAGTGGCAACCGCCAGAGTCCCTTTTA 1500
 |||||
 DB 2029 AGCAACGATTCATTCGATTCGATCAGAACACAGACAGTGGCAACCGCCAGAGTCCCTTTTA 2088
 |||||
 QY 1501 TCTCCACTGCATGGAAGTGGAGCGTGAAGCAATTACACACAGGCTTCTTTCCGGC 1560
 |||||

```
Db      2089  TCTCACTGCATCGAAGTGGAGGCTGGAGAGACATTACACACCAGCTTCCTTTGCGC 2148
QY      1561  CTTTCCACCCGACGACGACCTTCACGAAGTCTGCTGCGCTTAAAGGCTGGCAC 1620
Db      2149  CTTTCCACCGACGACGACCTTCACGAAGTCTGCTGCGCTTAAAGGCTGGCAC 2208
QY      1621  TCGGATATCTTGGCCCCCGACGACCTTCACCCCTTCCCTGACGACGCTGTATTGCGC 1680
Db      2209  TCGGATATCTTGGCCCCCGACGACCTTCACCCCTTCCCTGACGACGCTGTATTGCGC 2268
QY      1681  ACAGAGTCTCAACATTCTACTCTGCTCAGCCATCTACGAGGCAAGTCCAGTTACTCT 1740
Db      2269  ACAGAGTCTCAACATTCTACTCTGCTCAGCCATCTACGAGGCAAGTCCAGTTACTCT 2328
QY      1741  GCCTACAGCTGCAGCCGACGCTGCCCACTTGGGAGACCAAGTCTATTCTGTGCGCAGCGG 1800
Db      2329  GCCTACAGCTGCAGCCGACGCTGCCCACTTGGGAGACCAAGTCTATTCTGTGCGCAGCGG 2388
QY      1801  CAGAGCCAAAGTAGACAGAGCTGACTGCGCGCGAGCTGGCATGAGAGAGCCCTTTGAA 1860
Db      2389  CAGAGCCAAAGTAGACAGAGCTGACTGCGCGCGAGCTGGCATGAGAGAGCCCTTTGAA 2448
QY      1861  AAGCAGTTTAAACGACGAAGCTGCCAAATGGAATTTGAGAGACATCAATGTCAGAGAAC 1920
Db      2449  AAGCAGTTTAAACGACGAAGCTGCCAAATGGAATTTGAGAGACATCAATGTCAGAGAAC 2508
QY      1921  AGGTCACGGGAGAGAGCTGGGGAAAGTGGGACATCAAGCTTTTCGGGCAGCATGGA 1980
Db      2509  AGGTCACGGGAGAGAGCTGGGGAAAGTGGGACATCAAGCTTTTCGGGCAGCATGGA 2568
QY      1981  ATCATTGAGGTCTCC 1995
Db      2569  ATCATTGAGGTCTCC 2583
```

Search completed: June 22, 2004, 03:44:31
Job time : 1146 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2004, 03:25:27 ; Search time 3255 Seconds

(without alignments)
2807.764 Million cell updates/sec

Title: US-10-029-345a-108_COPY_538_2532

Perfect score: 1995
Sequence: 1 atggcccatgagatgattg.....tgggaatcatgaggtctcc 1995Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3017426 seqs, 2290544650 residues

Word size : 0

Total number of hits satisfying chosen parameters: 820462

Minimum DB seq length: 1419

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1893	94.9	1998	9 US-09-816-494-3	Sequence 3, Appl
2	1893	94.9	1998	16 US-10-377-072-27	Sequence 27, Appl
3	1893	94.9	3059	17 US-10-257-026-1	Sequence 1, Appl
4	1893	94.9	3496	9 US-09-964-277-1	Sequence 1, Appl
5	1893	94.9	3544	9 US-09-816-494-1	Sequence 1, Appl
6	1893	94.9	3544	16 US-10-377-072-25	Sequence 25, Appl
7	1893	94.9	3625	13 US-10-425-114-26234	Sequence 26234, A
8	1893	94.9	3766	13 US-10-343-357-117	Sequence 117, Appl
9	1893	94.9	4790	17 US-10-648-593-115	Sequence 115, Appl
10	1849	92.7	2966	13 US-10-296-115-520	Sequence 520, Appl
11	1842	92.3	2732	13 US-10-168-506-2	Sequence 2, Appl
12	1791	89.8	2102	16 US-10-094-749-673	Sequence 673, Appl
13	1418	71.1	3332	9 US-09-964-277-20	Sequence 20, Appl
14	1233	61.8	2071	13 US-10-072-012-257	Sequence 257, Appl

15	1182	59.2	2200	13 US-10-072-012-255	Sequence 255, Appl
16	788	39.5	1916	16 US-10-108-260A-2429	Sequence 2429, Appl
17	19	1.0	1449	12 US-10-052-482-192	Sequence 192, Appl
18	19	1.0	1635	17 US-10-648-593-27	Sequence 27, Appl
19	19	1.0	1830	15 US-10-346-356-1	Sequence 1, Appl
20	19	1.0	1909	12 US-10-052-482-191	Sequence 191, Appl
21	19	1.0	1909	13 US-10-058-270A-129	Sequence 129, Appl
22	19	1.0	2039	13 US-10-072-012-265	Sequence 265, Appl
23	19	1.0	2081	16 US-10-242-355-1076	Sequence 1076, Appl
24	19	1.0	2170	13 US-10-158-844-312	Sequence 312, Appl
25	19	1.0	2227	15 US-10-007-926A-330	Sequence 330, Appl
26	19	1.0	2227	16 US-10-435-696-54	Sequence 54, Appl
27	19	1.0	2387	16 US-10-443-694-100	Sequence 100, Appl
28	19	1.0	2476	13 US-10-220-120-28	Sequence 28, Appl
29	19	1.0	2631	17 US-10-378-448-3	Sequence 37, Appl
30	19	1.0	2653	9 US-09-969-708-536	Sequence 536, Appl
31	19	1.0	2653	10 US-09-960-706-716	Sequence 716, Appl
32	19	1.0	2653	10 US-09-873-319-451	Sequence 451, Appl
33	19	1.0	2653	15 US-10-094-699-2	Sequence 2, Appl
34	19	1.0	2653	15 US-10-205-823-131	Sequence 131, Appl
35	19	1.0	2653	15 US-10-210-120-5	Sequence 5, Appl
36	19	1.0	2653	16 US-10-341-434-70	Sequence 70, Appl
37	19	1.0	2653	16 US-10-117-937-7	Sequence 7, Appl
38	19	1.0	2653	16 US-10-443-694-1	Sequence 1, Appl
39	19	1.0	2655	15 US-10-252-157-9	Sequence 9, Appl
40	19	1.0	2731	16 US-10-264-048-686	Sequence 686, Appl
41	19	1.0	3017	16 US-10-443-694-38	Sequence 38, Appl
42	19	1.0	3017	16 US-10-443-694-128	Sequence 128, Appl
43	19	1.0	3077	16 US-10-443-694-90	Sequence 90, Appl
44	19	1.0	3286	16 US-09-971-392-9	Sequence 9, Appl
45	19	1.0	15500	9 US-09-764-869-2109	Sequence 2109, Appl

ALIGNMENTS

RESULT 1
US-09-816-494-3
; Sequence 3, Application US/09816494
; Patent No. US20020034807A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel A.
; TITLE OF INVENTION: 38692 AND 21117, NOVEL DUAL SPECIFICITY
; FILE REFERENCE: 10448-030002
; CURRENT APPLICATION NUMBER: US/09/816,494
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 60/191,858
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1998
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-816-494-3

Query Match	94.9%;	Score 1893;	DB 9;	Length 1998;
Best Local Similarity	99.9%;	Pred. No. 0;		
Matches 1993;	Conservative	0;	Mismatches 2;	Indels 0;
			Gaps	0;
QY	1	ATGGCCCATGAGATGATTGGAACTCAATTTGTTACTGAGAGTTGGTCTGCTGGGA	60	
DB	1	ATGGCCCATGAGATGATTGGAACTCAATTTGTTACTGAGAGTTGGTCTGCTGGGA	60	
QY	61	AGTGAACGGAAGAAAGTGTCTGATTTAGTACGGGCAATTTGTGGAATPACATATCC	120	
DB	61	AGTGAACGGAAGAAAGTGTCTGATTTAGTACGGGCAATTTGTGGAATPACATATCC	120	
QY	121	CACATTTTGAAGCCATTATATCACTCTCCACGTTATGAGGGAAGTTGCAACAG	180	
DB	121	CACATTTTGAAGCCATTATATCACTCTCCACGTTATGAGGGAAGTTGCAACAG	180	

PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 60/215,370
PRIOR FILING DATE: 2000-06-29
PRIOR APPLICATION NUMBER: US 09/723,806
PRIOR FILING DATE: 2000-11-28
PRIOR APPLICATION NUMBER: US 60/187,455
PRIOR FILING DATE: 2000-03-07
PRIOR APPLICATION NUMBER: US 09/843,297
PRIOR FILING DATE: 2001-04-25
PRIOR APPLICATION NUMBER: US 60/199,801
PRIOR FILING DATE: 2000-04-26
PRIOR APPLICATION NUMBER: US 09/861,801
PRIOR FILING DATE: 2001-05-21
PRIOR APPLICATION NUMBER: US 60/205,508
PRIOR FILING DATE: 2000-05-19
PRIOR APPLICATION NUMBER: US 09/816,494
PRIOR FILING DATE: 2001-03-23
PRIOR APPLICATION NUMBER: US 09/815,419
PRIOR FILING DATE: 2001-03-22
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 114
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 27
LENGTH: 1998
TYPE: DNA
ORGANISM: Homo Sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(1998)
US-10-377-072-27

Query Match 94.9%; Score 1893; DB 16; Length 1998;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCCCATGAGATGATGTTGGAATCAATATGTTACTGAGAGGTTGGTGGCTGTGCGAA 60
DB 1 ATGGCCCATGAGATGATGTTGGAATCAATATGTTACTGAGAGGTTGGTGGCTGTGCGAA 60
QY 61 AGTGAACGGAAAAAGTGTCTGCTAATTTGATAGCCGGCCATTTTGTGAAATACATATATCC 120
DB 61 AGTGAACGGAAAAAGTGTCTGCTAATTTGATAGCCGGCCATTTTGTGAAATACATATATCC 120
QY 121 CACATTTTGAACCATTAATATCACTGCTCAAGCTTATGAAAGGAAGTTGCAACG 180
DB 121 CACATTTTGAACCATTAATATCACTGCTCAAGCTTATGAAAGGAAGTTGCAACG 180
QY 181 GACAAAGTGTATTTACAGAGCTCATCCAGCATTTCAAGCGAAATTAAGTTGACATTGAT 240
DB 181 GACAAAGTGTATTTACAGAGCTCATCCAGCATTTCAAGCGAAATTAAGTTGACATTGAT 240
QY 241 TGCAGTCAGAAAGTTGATTTACGATCAAAAGCTCCAGAGTGTGCTCTCTCTTCA 300
DB 241 TGCAGTCAGAAAGTTGATTTACGATCAAAAGCTCCAGAGTGTGCTCTCTCTTCA 300
QY 301 GACTGTTTTCTCACTGACTTCTGGGTAACGAGAAAGCTTCAACTCTGTTCACTG 360
DB 301 GACTGTTTTCTCACTGACTTCTGGGTAACGAGAAAGCTTCAACTCTGTTCACTG 360
QY 361 CTGGCAGGTTGGGTTTGTGAGTCTCTCGTTGTTTCCCTGGCCCTGTGTAAGGAATAATCC 420
DB 361 CTGGCAGGTTGGGTTTGTGAGTCTCTCGTTGTTTCCCTGGCCCTGTGTAAGGAATAATCC 420
QY 421 ACTCTAGTCCCTACCTGATTTCTCAGCCTTGTGCTTACCTGTTGCAACATTTGGCCACCC 480
DB 421 ACTCTAGTCCCTACCTGATTTCTCAGCCTTGTGCTTACCTGTTGCAACATTTGGCCACCC 480
QY 481 CGAATTTCTCCCAATCTTTATCTTGGCTGCGAGGAGATGCTCTCAACAAAGAGCTGATG 540
DB 481 CGAATTTCTCCCAATCTTTATCTTGGCTGCGAGGAGATGCTCTCAACAAAGAGCTGATG 540
QY 541 CACAGAAATGGATGTGTTATGTTAAATGCCAGCTATACCTGTCCAAAGCCTGACTTT 600
DB 541 CACAGAAATGGATGTGTTATGTTAAATGCCAGCTATACCTGTCCAAAGCCTGACTTT 600

DB 541 CACAGAAATGGATGTGTTATGTTAAATGCCAGCAATACCTGTCCAAAGCCTGACTTT 600
QY 601 ATCCCGAATCTCATTTTCTGCGTGTGCTGGAATGACAGCTTTGTGAGAAATTTTG 660
DB 601 ATCCCGAATCTCATTTTCTGCGTGTGCTGGAATGACAGCTTTGTGAGAAATTTTG 660
QY 661 CCGTGTGAGCAATCACTAGATTTCAATTGAGAAAGCAAAAGCCTCAATGATGTGT 720
DB 661 CCGTGTGAGCAATCACTAGATTTCAATTGAGAAAGCAAAAGCCTCAATGATGTGT 720
QY 721 CTAGTCACTGTTTACCTGGGATCTCCGCTCCGCCACCATGCTATGCTTACATCATG 780
DB 721 CTAGTCACTGTTTACCTGGGATCTCCGCTCCGCCACCATGCTATGCTTACATCATG 780
QY 781 AAGAGATGACATGCTTTAATGATGAACTTACAGTTTGTGAAAGAAAAAGAACTTACT 840
DB 781 AAGAGATGACATGCTTTAATGATGAACTTACAGTTTGTGAAAGAAAAAGAACTTACT 840
QY 841 ATATCTCCAAACTTCAATTTTCTGGGCAACTCTGGAATATGAGAAAGATTAAGAAC 900
DB 841 ATATCTCCAAACTTCAATTTTCTGGGCAACTCTGGAATATGAGAAAGATTAAGAAC 900
QY 901 CAGACTGAGCATCAGGGCCAAAGCAAACTCAAGCTGTGCACTGAGAGAACCAAT 960
DB 901 CAGACTGAGCATCAGGGCCAAAGCAAACTCAAGCTGTGCACTGAGAGAACCAAT 960
QY 961 GAACTGTCTCTGCTGTCTCAGAGGTTGACAGAAAAGGAGAGCCCTCACTCAACC 1020
DB 961 GAACTGTCTCTGCTGTCTCAGAGGTTGACAGAAAAGGAGAGCCCTCACTCAACC 1020
QY 1021 TGTGCCAATCTGTACCTCAGAGGAGAGAGCAAAAGGCGGTGATCCCGCAGCGTG 1080
DB 1021 TGTGCCAATCTGTACCTCAGAGGAGAGAGCAAAAGGCGGTGATCCCGCAGCGTG 1080
QY 1081 CCCAGCGTCCAGAGCTGTGAGCGCTGTGTTAGAGAGAGCCGCTGTGATCAGCGCTTC 1140
DB 1081 CCCAGCGTCCAGAGCTGTGAGCGCTGTGTTAGAGAGAGCCGCTGTGATCAGCGCTTC 1140
QY 1141 AGTGGCTGCACTGTCTGCGAGACAGGCTGGAAGACCAATTAAGCTCAAGCTTCTTC 1200
DB 1141 AGTGGCTGCACTGTCTGCGAGACAGGCTGGAAGACCAATTAAGCTCAAGCTTCTTC 1200
QY 1201 TCTCTGGAATTAATCACTTTTCACTTATGAGCAGAGATGAGCATCTTACATGAGCTTC 1260
DB 1201 TCTCTGGAATTAATCACTTTTCACTTATGAGCAGAGATGAGCATCTTACATGAGCTTC 1260
QY 1261 TCTCATCAGAAAGATCTTTGGAATCTACAAACCTTCCACTACTGTGATGGAACCAAC 1320
DB 1261 TCTCATCAGAAAGATCTTTGGAATCTACAAACCTTCCACTACTGTGATGGAACCAAC 1320
QY 1321 AAGCTATGCAAGTCTCTCCCTGTTCAAGAACTATCCGAGCAGACTCCGAAACCAAGTCT 1380
DB 1321 AAGCTATGCAAGTCTCTCCCTGTTCAAGAACTATCCGAGCAGACTCCGAAACCAAGTCT 1380
QY 1381 GATTAAGAGAAAGCCAGCATCCCAAGAAAGCTGCAACCGCCAGGCTTCAAGCAGCAG 1440
DB 1381 GATTAAGAGAAAGCCAGCATCCCAAGAAAGCTGCAACCGCCAGGCTTCAAGCAGCAG 1440
QY 1441 AGCAAGCAATGCAATTCGCTCAGAAACCAAGCAGAGTGGCAACCGCCAGAGGTCCTTTTA 1500
DB 1441 AGCAAGCAATGCAATTCGCTCAGAAACCAAGCAGAGTGGCAACCGCCAGAGGTCCTTTTA 1500
QY 1501 TCTCACTGATGAAAGTGGAGAGCTGAGAGCAATTAACCAACAGCTTCTTTTGGC 1560
DB 1501 TCTCACTGATGAAAGTGGAGAGCTGAGAGCAATTAACCAACAGCTTCTTTTGGC 1560
QY 1561 CTTTCAACAGCAGAGAGCACTCAAGAGTCTGCTGGGCTTGAAGGCTGGCAC 1620
DB 1561 CTTTCAACAGCAGAGAGCACTCAAGAGTCTGCTGGGCTTGAAGGCTGGCAC 1620
QY 1621 TCGATATCTTGGCCCCCAGACCTTACACCTTCCCTGACAGAGCTGTATTTTGGC 1680
DB 1621 TCGATATCTTGGCCCCCAGACCTTACACCTTCCCTGACAGAGCTGTATTTTGGC 1680

QY 1681 ACAGAGTCTCA CACTTCTA CTGCTCAG CAGCATCTA CGAGGCA GTGCCAGTTACTCT 1740
|||
Db 1681 ACAGAGTCTCA CACTTCTA CTGCTCAG CAGCATCTA CGAGGCA GTGCCAGTTACTCT 1740
QY 1741 GCTTACAGCTG CAGGCACTG CCGCACTTGC GAGAGCA CCAAGTCTA TTTCTGTG CAGGCGG 1800
|||
Db 1741 GCTTACAGCTG CAGGCACTG CCGCACTTGC GAGAGCA CCAAGTCTA TTTCTGTG CAGGCGG 1800
QY 1801 CAGAGGCAAGTGA CAGAGCTGA CTGCGGCGG CAGCTGGCATG AAGAGAGCCCTTTGAA 1860
|||
Db 1801 CAGAGGCAAGTGA CAGAGCTGA CTGCGGCGG CAGCTGGCATG AAGAGAGCCCTTTGAA 1860
QY 1861 AAGCAGTTTAA CCGAGAGCTG CCAATGGA TTTGAGAGAGACATCATGTG CAGAGAAC 1920
|||
Db 1861 AAGCAGTTTAA CCGAGAGCTG CCAATGGA TTTGAGAGAGACATCATGTG CAGAGAAC 1920
QY 1921 AGGTCA CCGGAAAGCTG GGGGAAAGTG GGCAGTCA GTCTAGCTTTTTCGGG CAGCATGGA 1980
|||
Db 1921 AGGTCA CCGGAAAGCTG GGGGAAAGTG GGCAGTCA GTCTAGCTTTTTCGGG CAGCATGGA 1980
QY 1981 ATCATTTAGGTCTCC 1995
|||
Db 1981 ATCATTTAGGTCTCC 1995

RESULT 3

US-10-257-026-1
/ Sequence 1, Application US/10257026
/ Publication No. US20040086859A1
/ GENERAL INFORMATION:
/ APPLICANT: Merck Patent GmbH
/ TITLE OF INVENTION: New dual specificity phosphatase
/ FILE REFERENCE: DUSP10KXMS
/ CURRENT APPLICATION NUMBER: US/10/257,026
/ CURRENT FILING DATE: 2003-11-07
/ NUMBER OF SEQ. ID NOS: 3
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ. ID NO 1
/ LENGTH: 3059
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (127)..(2121)
US-10-257-026-1

Query Match 94.9%; Score 1893; DB 17; Length 3059;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCCCATGAGATGATTGGAATCTCAATTTGTTACTGAGAGGTGGTGGCTCTGCTGGAA 60
|||
Db 127 ATGGCCCATGAGATGATTGGAATCTCAATTTGTTACTGAGAGGTGGTGGCTCTGCTGGAA 186
QY 61 AATGGAAAGGAAAAAGTGTGCTTAATGATATGATACCGGCCATTGTTGGAATCAATATCATCC 120
|||
Db 187 AATGGAAAGGAAAAAGTGTGCTTAATGATATGATACCGGCCATTGTTGGAATCAATATCATCC 246
QY 121 CACATTTTGAAGCATTATATATCAACTGCTCCAGCTTTATGAGCGAAGGTGGAAAG 180
|||
Db 247 CACATTTTGAAGCATTATATATCAACTGCTCCAGCTTTATGAGCGAAGGTGGAAAG 306
QY 181 GACAAAGTTTATATACAGAGCTCATCCAGCATTCAGCGAAAATATAGGTTGACATTGAT 240
|||
Db 307 GACAAAGTTTATATACAGAGCTCATCCAGCATTCAGCGAAAATATAGGTTGACATTGAT 366
QY 241 TGCAGTGAAGAGTTGTGTTTGAATGATGAAAGCTCCCAAGATGTGGCTCTCTCTTCA 300
|||
Db 367 TGCAGTGAAGAGTTGTGTTTGAATGATGAAAGCTCCCAAGATGTGGCTCTCTCTTCA 426
QY 301 GACTGTTTCTCACTGTACTTCTGGGTAAACTGAGAGAGCTTCAACTGTCTCACTG 360
|||

Db 427 GACTGTTTCTCACTGTACTTCTGGGTAAACTGAGAGAGCTTCAACTGTGTCACTG 486
QY 361 CTTCGAGTGGGTTTCTGAGTTCTCTGGTTTCCCGGAGCTGTGGAAGAAAATCC 420
|||
Db 487 CTTCGAGTGGGTTTCTGAGTTCTCTGGTTTCCCGGAGCTGTGGAAGAAAATCC 546
QY 421 ACTCTAGTCCCTA CCGCATTTTCTGACGCTTCTTA CCGTTTGCCAACAATTGGGCAAC 480
|||
Db 547 ACTCTAGTCCCTA CCGCATTTTCTGACGCTTCTTA CCGTTTGCCAACAATTGGGCAAC 606
QY 481 CGAATTTCTCCCAATCTTTATCTTGCGCCAGCGAGATGCTCAACAAGAGCTGTATA 540
|||
Db 607 CGAATTTCTCCCAATCTTTATCTTGCGCCAGCGAGATGCTCAACAAGAGCTGTATA 666
QY 541 CAGCAAAATGGAGTTGTTATGTTTAAATGCAAGCTATA CCGTCCAAAGCTGACTT 600
|||
Db 667 CAGCAAAATGGAGTTGTTATGTTTAAATGCAAGCTATA CCGTCCAAAGCTGACTT 726
QY 601 ATCCCCGAGTCAATTTCTGCGTGTGCTGTGAATGACAGCTTTTGTGAGAAAATTTTG 660
|||
Db 727 ATCCCCGAGTCAATTTCTGCGTGTGCTGTGAATGACAGCTTTTGTGAGAAAATTTTG 786
QY 661 CCGTGTGGA CAAATCAGTAGATTTCATGAGAAAGCAAAAGCTCCAAATGATGTGT 720
|||
Db 787 CCGTGTGGA CAAATCAGTAGATTTCATGAGAAAGCAAAAGCTCCAAATGATGTGT 846
QY 721 CTAGTGCACTGTTTACGTGGAGATCTCCGCTCCGCAACCATGCTATGCTTCATCATATG 780
|||
Db 847 CTAGTGCACTGTTTACGTGGAGATCTCCGCTCCGCAACCATGCTATGCTTCATCATATG 906
QY 781 AAGAGATGGA CAGTCTTTTATGATGAGCTTACAGATTGTTGAAAAGAAAAGCTACT 840
|||
Db 907 AAGAGATGGA CAGTCTTTTATGATGAGCTTACAGATTGTTGAAAAGAAAAGCTACT 966
QY 841 ATATCTCCAACTTCAATTTTCTGGGCCAACTCCGTGACTATGAGAGAAATTAAGAAC 900
|||
Db 967 ATATCTCCAACTTCAATTTTCTGGGCCAACTCCGTGACTATGAGAGAAATTAAGAAC 1026
QY 901 CAGACTGGAAGATCAGAGGCCCAAAAGCAAACTCAAGCTGCTGCA CTTGAGAAAGCCAAAT 960
|||
Db 1027 CAGACTGGAAGATCAGAGGCCCAAAAGCAAACTCAAGCTGCTGCA CTTGAGAAAGCCAAAT 1086
QY 961 GAACTGTGCTGTGCTCAGAGAGGTGGA CAGAAAAGCGAGAGCCCTCACTCCACCC 1020
|||
Db 1087 GAACTGTGCTGTGCTCAGAGAGGTGGA CAGAAAAGCGAGAGCCCTCACTCCACCC 1146
QY 1021 TGTGCCAATCTGTGCTACCTCAGAGGCAAGGCAAAAGCCGTGATCCCGCAGCGTG 1080
|||
Db 1147 TGTGCCAATCTGTGCTACCTCAGAGGCAAGGCAAAAGCCGTGATCCCGCAGCGTG 1206
QY 1081 CCGAGCGTCCAGCGGTGAGCCGTGCTGTTTGAAGACAGCCCGTGTGATCAGGCGCTC 1140
|||
Db 1207 CCGAGCGTCCAGCGGTGAGCCGTGCTGTTTGAAGACAGCCCGTGTGATCAGGCGCTC 1266
QY 1141 AATGGGCTGCACTGTGCTCCGAGACAGGCTGGAAGACAGCAATATAGCTCAAGCTTCTTC 1200
|||
Db 1267 AATGGGCTGCACTGTGCTCCGAGACAGGCTGGAAGACAGCAATATAGCTCAAGCTTCTTC 1326
QY 1201 TCTCTGATATCAAAATCATGATTCAATATTCAGCCAGATGGCAGACATCTTAATGAGCTTC 1260
|||
Db 1327 TCTCTGATATCAAAATCATGATTCAATATTCAGCCAGATGGCAGACATCTTAATGAGCTTC 1386
QY 1261 TCTCTCATGAGAAATGCTTTGGAATCTTCAAACTTTCATCTGTGATGGGACCAAC 1320
|||
Db 1387 TCTCTCATGAGAAATGCTTTGGAATCTTCAAACTTTCATCTGTGATGGGACCAAC 1446
QY 1321 AAGCTATGCCAGTTCTCCCTGTGTCAGAACTATGGAACACTCCGAAACCAAGTCTT 1380
|||
Db 1447 AAGCTATGCCAGTTCTCCCTGTGTCAGAACTATGGAACACTCCGAAACCAAGTCTT 1506
QY 1381 GATTAAGAGAAAGCCAGCATTCGCCAAAGACTGTCAGACCGCCAGAGCTTTCAGACGCCAG 1440
|||
Db 1507 GATTAAGAGAAAGCCAGCATTCGCCAAAGACTGTCAGACCGCCAGAGCTTTCAGACGCCAG 1566

QY 1441 AGCAAGCGATTGGATTGGTCAGAACACAGAGGTCGACCGCCAGAGGTCCCTTTA 1500
| | | | |
Db 1567 AGCAAGCGATTGGATTGGTCAGAACACAGAGGTCGACCGCCAGAGGTCCCTTTA 1626
QY 1501 TCTCCACTGCATCGAAGTGGAGCGTCGAGAGCAATTAACACACAGTTCCTTTTCGCG 1560
| | | | |
Db 1627 TCTCCACTGCATCGAAGTGGAGCGTCGAGAGCAATTAACACACAGTTCCTTTTCGCG 1686
QY 1561 CTTTCCACAGCAGCAGACAGCCTCAGCAAGTCTGTCGGCCTTGAAGGGCTGGCAC 1620
| | | | |
Db 1687 CTTTCCACAGCAGCAGACAGCCTCAGCAAGTCTGTCGGCCTTGAAGGGCTGGCAC 1746
QY 1621 TCGGATATCTGGCCCCCGCCAGACCTTACCCCTTCCCTGACAGCAGCTGTATTTTGGC 1680
| | | | |
Db 1747 TCGGATATCTGGCCCCCGCCAGACCTTACCCCTTCCCTGACAGCAGCTGTATTTTGGC 1806
QY 1681 ACAGAGTCCCTACACTTCTACTGCTGCTGACGACCTTACAGAGGTCAGTTACTCT 1740
| | | | |
Db 1807 ACAGAGTCCCTACACTTCTACTGCTGCTGACGACCTTACAGAGGTCAGTTACTCT 1866
QY 1741 GCCTACAGCTGACGACGCTGCCACTTGGCGAGACCAAGTCTATTTCTGTGGCAGCGG 1800
| | | | |
Db 1867 GCCTACAGCTGACGACGCTGCCACTTGGCGAGACCAAGTCTATTTCTGTGGCAGCGG 1926
QY 1801 CAGAACCCAGTACAGAGCTGACTGCTGGCGCGAGCTGGCATGAAAGAGCCCTTTTGA 1860
| | | | |
Db 1927 CAGAACCCAGTACAGAGCTGACTGCTGGCGCGAGCTGGCATGAAAGAGCCCTTTTGA 1986
QY 1861 AAGCAGTTTAAACGAGAGCTGCCAAATGGAATTTGAGAGAGCATCTGTCAAGAAC 1920
| | | | |
Db 1987 AAGCAGTTTAAACGAGAGCTGCCAAATGGAATTTGAGAGAGCATCTGTCAAGAAC 2046
QY 1921 AAGTCACGGAAAGAGTGGGAAAGTGGGAGTCACTTTTGGGCGAGCATGAA 1980
| | | | |
Db 2047 AAGTCACGGAAAGAGTGGGAAAGTGGGAGTCACTTTTGGGCGAGCATGAA 2106
QY 1981 ATCATTGAGTCTCC 1995
| | | | |
Db 2107 ATCATTGAGTCTCC 2121
| | | | |
RESULT 4
US-09-964-277-1
; Sequence 1, Application US/09964277
; Patent No. US20020137170A1
; GENERAL INFORMATION:
; APPLICANT: Lucbe, Ralf M.
; TITLE OF INVENTION: DSP-16 DUAL-SPECIFICITY PHOSPHATASE
; FILE REFERENCE: 200125.434
; CURRENT APPLICATION NUMBER: US/09/964,277
; CURRENT FILING DATE: 2001-09-25
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3496
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-964-277-1
Query Match 94.9%; Score 1893; DB 9; Length 3496;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 121 CACATTTTGAAGCATTAAATATCAATGCTCCCAAGCTTATGAGCGAAGTTGCAACAG 180
| | | | |
Db 682 CACATTTTGAAGCATTAAATATCAATGCTCCCAAGCTTATGAGCGAAGTTGCAACAG 741
QY 181 GACAAAGTTTAAATTAACAGAGCTCATCAAGATTCAAGCGAAACATTAAGTTGACATTGAT 240
| | | | |
Db 742 GACAAAGTTTAAATTAACAGAGCTCATCAAGATTCAAGCGAAACATTAAGTTGACATTGAT 801
QY 241 TGCAGTCAAGAGTTGATTTAGATCAAAAGCTCCCAAGATGGTCCCTCTCTCTTCA 300
| | | | |
Db 802 TGCAGTCAAGAGTTGATTTAGATCAAAAGCTCCCAAGATGGTCCCTCTCTCTTCA 861
QY 301 GACTGTTTCTCACTGTAATCTTGGGTAACTGAGAAAGAGCTTCAACTCTGTCACCTG 360
| | | | |
Db 862 GACTGTTTCTCACTGTAATCTTGGGTAACTGAGAAAGAGCTTCAACTCTGTCACCTG 921
QY 361 CTGCAAGTGGGTTTGTGAGTCTCTGCTGTTTCCCTGGCTCTGTGAAGAAATCC 420
| | | | |
Db 922 CTGCAAGTGGGTTTGTGAGTCTCTGCTGTTTCCCTGGCTCTGTGAAGAAATCC 981
QY 421 ACTCTAGTCCCTACCTGCAATTTCTAGCTTCTGCTTACCTGTTGGCCAACTTGGGCCAAC 480
| | | | |
Db 982 ACTCTAGTCCCTACCTGCAATTTCTAGCTTCTGCTTACCTGTTGGCCAACTTGGGCCAAC 1041
QY 481 CGAATTTCTTCCCAATCTTTATCTTGGGTGCGCAGCGAGATGTCCTCAACAGAGAGCTGATA 540
| | | | |
Db 1042 CGAATTTCTTCCCAATCTTTATCTTGGGTGCGCAGCGAGATGTCCTCAACAGAGAGCTGATA 1101
QY 541 CAGCAAAATGGGATTGGTTATGTGTTAAATGCGAGCTATACCTGTCCAAAGCTGACTG 600
| | | | |
Db 1102 CAGCAAAATGGGATTGGTTATGTGTTAAATGCGAGCTATACCTGTCCAAAGCTGACTG 1161
QY 601 ATCCCGAGTCTCAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660
| | | | |
Db 1162 ATCCCGAGTCTCAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1221
QY 661 CCGTGGTTGGAACAAATCAGTATGATTTCAATGAGAAAGCAAAAGCTCCAAATGATGTT 720
| | | | |
Db 1222 CCGTGGTTGGAACAAATCAGTATGATTTCAATGAGAAAGCAAAAGCTCCAAATGATGTT 1281
QY 721 CTATGCACTGTTTATGCTGGGATCTCCGCTCCGCAACCATGCTATGCTTACATCATG 780
| | | | |
Db 1282 CTATGCACTGTTTATGCTGGGATCTCCGCTCCGCAACCATGCTATGCTTACATCATG 1341
QY 781 AAGAGATGACATGCTTTTATGATGAAGTTTACAGATTTTGTAAAGAAAGAAAGAAAGCTTACT 840
| | | | |
Db 1342 AAGAGATGACATGCTTTTATGATGAAGTTTACAGATTTTGTAAAGAAAGAAAGAAAGCTTACT 1401
QY 841 ATATCTCCAACTTCAATTTTCTGGGCCAACTCTGGAATATGAGAAAGATTAAGAC 900
| | | | |
Db 1402 ATATCTCCAACTTCAATTTTCTGGGCCAACTCTGGAATATGAGAAAGATTAAGAC 1461
QY 901 CAGACTGAGCATCAGAGGCGCAAGAGCAAACTCAAGCTGCTCAGCTGAGAGAGCCCAAT 960
| | | | |
Db 1462 CAGACTGAGCATCAGAGGCGCAAGAGCAAACTCAAGCTGCTCAGCTGAGAGAGCCCAAT 1521
QY 961 GAACTGTCTCTGCTGTCTCAAGAGGTGACAGAAAGAGCAAGCCCTCTCACTCAACC 1020
| | | | |
Db 1522 GAACTGTCTCTGCTGTCTCAAGAGGTGACAGAAAGAGCAAGCCCTCTCACTCAACC 1581
QY 1021 TGTGCGGATCTGCTACCTCAGAGGACAGAGCAAAAGCCGCTGATCCCGCAGCGTG 1080
| | | | |
Db 1582 TGTGCGGATCTGCTACCTCAGAGGACAGAGCAAAAGCCGCTGATCCCGCAGCGTG 1641
QY 1081 CCCAGGTCGCCAGAGGTGAGCGCTGCTGTTAGAGACAGCCCGCTGTACAGGGGCTC 1140
| | | | |
Db 1642 CCCAGGTCGCCAGAGGTGAGCGCTGCTGTTAGAGACAGCCCGCTGTACAGGGGCTC 1701
QY 1141 AGTGGCTGCACTGTCCGCAACAGGCTGAAAGACAGCAATTAAGTCAAGGTTCTTC 1200
| | | | |
Db 1702 AGTGGCTGCACTGTCCGCAACAGGCTGAAAGACAGCAATTAAGTCAAGGTTCTTC 1761
QY 1201 TCTCTGATATCAATCACTTCAATTTCAAGCAGATGCGAGCATCTTACATGCTTC 1260
| | | | |

```
Db 1762 TCTCTGATATCAAAATCAGTTTCATATTCAGCCGAGATGAGATCTTACATGGCTTC 1821
Qy 1261 TCCCTCAGAGAAAGTGTGGTGGAACTACTACAAACCTTCCACTACTCTGGATGGAGCCAC 1320
Db 1822 TCCCTCAGAGAAAGTGTGGTGGAACTACTACAAACCTTCCACTACTCTGGATGGAGCCAC 1881
Qy 1321 AAGCTATGCCAGTCTTCCCTGTTCCAGAACTATCGAGAGAGACTCCCGAAACAGTCTC 1380
Db 1882 AAGCTATGCCAGTCTTCCCTGTTCCAGAACTATCGAGAGAGACTCCCGAAACAGTCTC 1941
Qy 1381 GATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
Db 1942 GATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2001
Qy 1441 AGAAGAGATTTGATTTGGTCTGATCAAGACAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
Db 2002 AGAAGAGATTTGATTTGGTCTGATCAAGACAGAGAGAGAGAGAGAGAGAGAGAGAG 2061
Qy 1501 TCTCACTGATCGAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
Db 2062 TCTCACTGATCGAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2121
Qy 1561 CTTTCCACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
Db 2122 CTTTCCACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2181
Qy 1621 TCGGATATCTTGGCCCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
Db 2182 TCGGATATCTTGGCCCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2241
Qy 1681 ACAGAGTCTCAGACTTCTACTCTGCTCAGCCATCAGAGAGAGAGAGAGAGAGAGAG 1740
Db 2242 ACAGAGTCTCAGACTTCTACTCTGCTCAGCCATCAGAGAGAGAGAGAGAGAGAGAG 2301
Qy 1741 GCTTAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1800
Db 2302 GCTTAAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2361
Qy 1801 CAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
Db 2362 CAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2421
Qy 1861 AAGCAGTTTAAAGCAGAAAGCTGCCAAATGGAATTTGAGAGAGAGAGAGAGAGAG 1920
Db 2422 AAGCAGTTTAAAGCAGAAAGCTGCCAAATGGAATTTGAGAGAGAGAGAGAGAGAG 2481
Qy 1921 AAGTACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
Db 2482 AAGTACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2541
Qy 1981 ATCATTGAGTCTCC 1995
Db 2542 ATCATTGAGTCTCC 2556

RESULT 5
US-09-816-494-1
; Sequence 1, Application US/09816494
; Patent No. US20020034807A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel A.
; TITLE OF INVENTION: 38692 AND 21117, NOVEL DUAL SPECIFICITY
; FILE REFERENCE: 10448-030002
; CURRENT APPLICATION NUMBER: US/09/816,494
; PRIOR APPLICATION NUMBER: US 60/191,858
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3544
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (589) ... (2583)
US-09-816-494-1

Query Match          94.9%; Score 1893; DB 9; Length 3544;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGCCCATGAGATGATTTGAACTCAATTTGTTACTGAGAGAGTGTGGCTCTGCTGAA 60
Db 589 ATGGCCCATGAGATGATTTGAACTCAATTTGTTACTGAGAGAGTGTGGCTCTGCTGAA 648
Qy 61 AGTGAAGGAAAAAGTGTCTGTAATGATAGCCGCAATTTGGAAATACAAATCATCC 120
Db 649 AGTGAAGGAAAAAGTGTCTGTAATGATAGCCGCAATTTGGAAATACAAATCATCC 708
Qy 121 CACATTTTGAAGCATTAATATCACTGCTCCAGCTTAATGAAGCGAAAGTTGCAACG 180
Db 709 CACATTTTGAAGCATTAATATCACTGCTCCAGCTTAATGAAGCGAAAGTTGCAACG 768
Qy 181 GACAAAGTGAATTAATACAGAGCTCATCCAGATTACAGGAAACATAAGATTGACATTGAT 240
Db 769 GACAAAGTGAATTAATACAGAGCTCATCCAGATTACAGGAAACATAAGATTGACATTGAT 828
Qy 241 TGCAGTCAAGAGTGTGATTTAGATCAAAAGCTCCCAAGATGTTCCTCTCTCTCA 300
Db 829 TGCAGTCAAGAGTGTGATTTAGATCAAAAGCTCCCAAGATGTTCCTCTCTCTCTCA 888
Qy 301 GACTGTTTCTCAGTCACTTCTGGGTAACTGGAGAAAGAGCTCAACTCTGTTCACTG 360
Db 889 GACTGTTTCTCAGTCACTTCTGGGTAACTGGAGAAAGAGCTCAACTCTGTTCACTG 948
Qy 361 CTTGAGAGTGGATTTGCTGAGATTCTCTGTTGTTTCCCTGAGCTCTGTGAAGAAATCC 420
Db 949 CTTGAGAGTGGATTTGCTGAGATTCTCTGTTGTTTCCCTGAGCTCTGTGAAGAAATCC 1008
Qy 421 ACTCAAGTCTTACCTGATTTCTCAGAGCTTGTCTTACTCTGTGCCAACATTGGGCCAAC 480
Db 1009 ACTCAAGTCTTACCTGATTTCTCAGAGCTTGTCTTACTCTGTGCCAACATTGGGCCAAC 1068
Qy 481 GCAATTCCTCCCAATCTTATCTTGGCTGCGCAGAGAGATGCTCAACAGAGAGCTGATA 540
Db 1069 GCAATTCCTCCCAATCTTATCTTGGCTGCGCAGAGAGATGCTCAACAGAGAGCTGATA 1128
Qy 541 CAGAGAAATGGGATTTGTTATGTTAAATGCAAGCTTAACTGTCCAAAGCCTGACTTT 600
Db 1129 CAGAGAAATGGGATTTGTTATGTTAAATGCAAGCTTAACTGTCCAAAGCCTGACTTT 1188
Qy 601 ATCCCGAGTCTCAATTCCTGCGGTGCTGTGTAATGCAAGAGCTTTTGTGAAGAAATTTTG 660
Db 1189 ATCCCGAGTCTCAATTCCTGCGGTGCTGTGTAATGCAAGAGCTTTTGTGAAGAAATTTTG 1248
Qy 661 CCGTGTGTTGCAAAATCAGTATGTTTCAATTTGAGAAAGCAAAAGCCCAATGATGTG 720
Db 1249 CCGTGTGTTGCAAAATCAGTATGTTTCAATTTGAGAAAGCAAAAGCCCAATGATGTG 1308
Qy 721 CTATGCACTGTTTAAAGTGGATCTCCGCTCCGCAACATCGTATCGCTTACATCATG 780
Db 1309 CTATGCACTGTTTAAAGTGGATCTCCGCTCCGCAACATCGTATCGCTTACATCATG 1368
Qy 781 AAGAGATGAGACATGCTTTTATGATGAGAGCTTACAGATTTGTGAAGAAAGAAAGCTACT 840
Db 1369 AAGAGATGAGACATGCTTTTATGATGAGAGCTTACAGATTTGTGAAGAAAGAAAGCTACT 1428
Qy 841 ATATCTCAAACTTCAATTTTCTGGGCCAATCTCTGAGATATGAGAAAGAGATTTAAGAAC 900
Db 1429 ATATCTCAAACTTCAATTTTCTGGGCCAATCTCTGAGATATGAGAAAGAGATTTAAGAAC 1488
Qy 901 CAGACTGAGCATCAGGCGCAAAAGAGCAAACTCAAGCTGTGCACTGAGAAAGCAAT 960
```


241 TCGAGTCAGAAAGTTGTAAGTTTACGATCAAGGCTCCCAAGATGTTGCTCTCTCTTCA 300
829 TCGAGTCAGAAAGTTGTAAGTTTACGATCAAGGCTCCCAAGATGTTGCTCTCTCTTCA 888
301 GACTGTTTTCTCACTGTAATCTTGGGTAACTGGAGAGAAGCTTCACTCTGTTACCTG 360
889 GACTGTTTTCTCACTGTAATCTTGGGTAACTGGAGAGAAGCTTCACTCTGTTACCTG 948
361 CTTCGAGGTGGGTTTCTGAGTTCTCTGTTGTTTCCCTGGCTCTGTGAAGAAATTC 420
949 CTTCGAGGTGGGTTTCTGAGTTCTCTGTTGTTTCCCTGGCTCTGTGAAGAAATTC 1008
421 ACTCTAGTCCCTACCTGCAATTTCTGAGCTTGTCTGTTACCTGTTGCAACATTTGGGCAAC 480
1009 ACTCTAGTCCCTACCTGCAATTTCTGAGCTTGTCTGTTACCTGTTGCAACATTTGGGCAAC 1068
481 CGAATTTCTTCCCAATCTTTATCTTGGCTGCAAGCGATGCTCTCAACAAGAGCTGATA 540
1069 CGAATTTCTTCCCAATCTTTATCTTGGCTGCAAGCGATGCTCTCAACAAGAGCTGATA 1128
541 CAGCAAAATGGGATTTGTTATGTTTAAATGCAAGTAACTCTGTTCAAAAGCTGACTT 600
1129 CAGCAAAATGGGATTTGTTATGTTTAAATGCAAGTAACTCTGTTCAAAAGCTGACTT 1188
601 ATCCCGAGTCAATTTCCGCGTGTGCTGATGATGACAGCTTTTGTGAAAAATTTTG 660
1189 ATCCCGAGTCAATTTCCGCGTGTGCTGATGATGACAGCTTTTGTGAAAAATTTTG 1248
661 CCGTGGTTGCAAAATCACTAGATTTCAATTGAGAAAAAGAAAGCTTCAATGATGTGT 720
1249 CCGTGGTTGCAAAATCACTAGATTTCAATTGAGAAAAAGAAAGCTTCAATGATGTGT 1308
721 CTAAGTCACTGTTTGTGCTGGGATCTCCGCTCCGCAACATGCTATGCTTCAATCANG 780
1309 CTAAGTCACTGTTTGTGCTGGGATCTCCGCTCCGCAACATGCTATGCTTCAATCANG 1368
781 AAGAGATGAGATGCTTTAGATGAGCTTACAGATTTGTGAAAAAGAAAGCTTCACT 840
1369 AAGAGATGAGATGCTTTAGATGAGCTTACAGATTTGTGAAAAAGAAAGCTTCACT 1428
841 AATATCCAACTTCAATTTTCTGGGCAACTCTCTGACTATGAGAAAGATTAAGAAC 900
1429 AATATCCAACTTCAATTTTCTGGGCAACTCTCTGACTATGAGAAAGATTAAGAAC 1488
901 CAGACTGAGCATCAGGGGCAAAAGCAAACTCAAGCTGCTCAGCTGAGAAAGCTCAAT 960
1489 CAGACTGAGCATCAGGGGCAAAAGCAAACTCAAGCTGCTCAGCTGAGAAAGCTCAAT 1548
961 GAACCTGCTCCTGCTGTCTCAGAGGGTGACAGAAAAGGAGAGCGCCCTCAGTCCACC 1020
1549 GAACCTGCTCCTGCTGTCTCAGAGGGTGACAGAAAAGGAGAGCGCCCTCAGTCCACC 1608
1021 TGTGCCAATCTGCTACTCAAGAGCAGAGCAAAAGGCGCTGATCCGCCACGCTG 1080
1609 TGTGCCAATCTGCTACTCAAGAGCAGAGCAAAAGGCGCTGATCCGCCACGCTG 1668
1081 CCCAGGTCAGAGGTGAGCGCTGCTGTAGAGGACAGCGCTGTTAAGCGCTC 1140
1669 CCCAGGTCAGAGGTGAGCGCTGCTGTAGAGGACAGCGCTGTTAAGCGCTC 1728
1141 AGTGGGTCGACCTGTCCGACAGAGGCTGGAAGACAGCAATTAAGCTTCAAGCTTCTTC 1200
1729 AGTGGGTCGACCTGTCCGACAGAGGCTGGAAGACAGCAATTAAGCTTCAAGCTTCTTC 1788
1201 TCTCTGATATCAATCAAGTTTCAATTTCAAGCAGATGAGCATCTTACATGAGCTTC 1260
1789 TCTCTGATATCAATCAAGTTTCAATTTCAAGCAGATGAGCATCTTACATGAGCTTC 1848
1261 TCCATATCAAGAGATGCTTTGGAATTAATAAAGCTTCACTACTCTGATGAGCAAC 1320
1849 TCCATATCAAGAGATGCTTTGGAATTAATAAAGCTTCACTACTCTGATGAGCAAC 1908

QY 1321 AAGCTATGCAAGTTTCTCCCTGTTCTAGGAACTATCGAGAGACAGATCTCCGAAACAGTCT 1380
DB 1909 AAGCTATGCAAGTTTCTCCCTGTTCTAGGAACTATCGAGAGACAGATCTCCGAAACAGTCT 1968
QY 1381 GATTAAGAGAAAGCCAGCATCTCCCAAGAGCTGACAGCCGCCAGGCTTTACAGACAGCAG 1440
DB 1969 GATTAAGAGAAAGCCAGCATCTCCCAAGAGCTGACAGCCGCCAGGCTTTACAGACAGCAG 2028
QY 1441 AGCAAGCATTTGCAATTTGGTCTAGAAACAGACAGCATGAGCACCGCCAGAGTCCCTTTTA 1500
DB 2029 AGCAAGCATTTGCAATTTGGTCTAGAAACAGACAGCATGAGCACCGCCAGAGTCCCTTTTA 2088
QY 1501 TCTCCATGTCATTCGAAGTGGAGAGCTGGAGAGCAATTAACACACAGCTTCTTTTGGC 1560
DB 2089 TCTCCATGTCATTCGAAGTGGAGAGCTGGAGAGCAATTAACACACAGCTTCTTTTGGC 2148
QY 1561 CTTTCACACAGCAGACACCTCAAGAGTCTGAGCTTGGGCTTTAAGGCTGGCAG 1620
DB 2149 CTTTCACACAGCAGACACCTCAAGAGTCTGAGCTTGGGCTTTAAGGCTGGCAG 2208
QY 1621 TCGATATCTTGGGCTCCCAAGCTTACCTTCCCTGACACAGAGCTGATTTTGGC 1680
DB 2209 TCGATATCTTGGGCTCCCAAGCTTACCTTCCCTGACACAGAGCTGATTTTGGC 2268
QY 1681 ACAGAGCTCTCACTTCTACTGCTGCTCAGCATCTACAGAGCAGTGCAGTACTCT 1740
DB 2269 ACAGAGCTCTCACTTCTACTGCTGCTCAGCATCTACAGAGCAGTGCAGTACTCT 2328
QY 1741 GCTTACAGCTGAGCAGCTGCTGCCAATTGGGAGACCAAGTCTATTCTGTGCGCAGCGG 1800
DB 2329 GCTTACAGCTGAGCAGCTGCTGCCAATTGGGAGACCAAGTCTATTCTGTGCGCAGCGG 2388
QY 1801 CAGAAAGCAAGTGCAGAGCTGACTCGCGGGGAGAGCTGGCAATGAAGAGAGCCCTTTGA 1860
DB 2389 CAGAAAGCAAGTGCAGAGCTGACTCGCGGGGAGAGCTGGCAATGAAGAGAGCCCTTTGA 2448
QY 1861 AAGCATTTAAACGCAAGCTGCCAATTTGAGAGAGAGCATGATCAGAGAAC 1920
DB 2449 AAGCATTTAAACGCAAGCTGCCAATTTGAGAGAGAGCATGATCAGAGAAC 2508
QY 1921 AGTTCACGGAAAGAGCTGGGAAAGTGGGAGTCACTTACCTTTTCCGGCAGATGGA 1980
DB 2509 AGTTCACGGAAAGAGCTGGGAAAGTGGGAGTCACTTACCTTTTCCGGCAGATGGA 2568
QY 1981 ATCAATTGAGTCTCC 1995
DB 2569 ATCAATTGAGTCTCC 2583

RESULT 7
US-10-425-114-26234

Sequence 26234, Application US/10425114
Publication No. US20040034888A1
GENERAL INFORMATION:
APPLICANT: Liu, Jindong
APPLICANT: Zhou, Yihua
APPLICANT: Kowalic, David K.
APPLICANT: Screen, Steven E.
APPLICANT: Tabaska, Jack E.
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
FILE REFERENCE: 38-21(5313)B
CURRENT APPLICATION NUMBER: US/10/425,114
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 26234
LENGTH: 3625
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Clone ID: LIB4119-028-H6_FLI
US-10-425-114-26234


```
GENERAL INFORMATION:
APPLICANT: INCYTE GENOMICS, INC.; TANG, Y.Tom
APPLICANT: ELLIOTT, Vicki S.; RAMKUMAR, Jayalaxmi
APPLICANT: YAO, Monique G.; BURFORD, Neil
APPLICANT: WANG, Yumei E.; STEWART, Elizabeth A.
APPLICANT: GANDHI, Ameena R.; ARVIZU, Chandra S.
APPLICANT: LEE, Ernestine A.; HAFALIA, April J.A.
APPLICANT: LU, Dying Alina M.; TRIBOULEY, Catherine M.
APPLICANT: GRIFFIN, Jennifer A.; BAUGHN, Mariah R.
APPLICANT: YUE, Henry; WARREN, Bridget A.
APPLICANT: NGUYEN, Daniel B.; CHAMLA, Narinder K.
APPLICANT: KEARNEY, Liam
TITLE OF INVENTION: PROTEIN PHOSPHATASES
FILE REFERENCE: PI-0173 PCT
CURRENT APPLICATION NUMBER: US/10/343,357
CURRENT FILING DATE: 2003-01-28
PRIOR APPLICATION NUMBER: PCT/US01/23716
PRIOR FILING DATE: 2001-07-26
PRIOR APPLICATION NUMBER: US 60/221,679
PRIOR FILING DATE: 2000-07-28
PRIOR APPLICATION NUMBER: US 60/223,272
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: US 60/224,309
PRIOR FILING DATE: 2000-08-10
PRIOR APPLICATION NUMBER: US 60/226,728
PRIOR FILING DATE: 2000-08-18
PRIOR APPLICATION NUMBER: US 60/229,254
PRIOR FILING DATE: 2000-08-30
PRIOR APPLICATION NUMBER: US 60/231,366
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 20
SOFTWARE: PERL Program
SEQ ID NO 17
LENGTH: 3766
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: Incyte ID No. US20040058341A1 7480570CB1
US-10-343-357-17

Query Match      94.9%; Score 1893; DB 13; Length 3766;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 ATGGCCCATGATGATGATGAACTCAATTTGTTACTGAGAGGTGGTGGCTGTGCGAA 60
DB      538 ATGGCCCATGATGATGATGAACTCAATTTGTTACTGAGAGGTGGTGGCTGTGCGAA 597
QY      61 AGTGAACGGAAAAAGTGTGCTTAATTGATAGCCGGCCATTGTTGGAATACAAATCATCC 120
DB      598 AGTGAACGGAAAAAGTGTGCTTAATTGATAGCCGGCCATTGTTGGAATACAAATCATCC 657
QY      121 CACATTTTGGAGCCATTATATCACTGCTCCAAAGCTTATGAAAGGAAGTTGCAACAG 180
DB      658 CACATTTTGGAGCCATTATATCACTGCTCCAAAGCTTATGAAAGGAAGTTGCAACAG 717
QY      181 GACAAAGTTTATTACAGAGCTCATCCAGATTCAGAGGAACATTAAGTTGACATTAT 240
DB      718 GACAAAGTTTATTACAGAGCTCATCCAGATTCAGAGGAACATTAAGTTGACATTAT 777
QY      241 TGCAGTCAGAAAGTTGATTTTATGATCAAAAGCTCCCAAGATGTTGCTCTCTCTCA 300
DB      778 TGCAGTCAGAAAGTTGATTTTATGATCAAAAGCTCCCAAGATGTTGCTCTCTCTCA 837
QY      301 GACTTTTTTTCACGTACTTCTGGGTAACTGGGAAAGAGCTTAACTCTGTTCACCTG 360
DB      838 GACTTTTTTTCACGTACTTCTGGGTAACTGGGAAAGAGCTTAACTCTGTTCACCTG 897
QY      361 CTTCAGAGTGGGTTTGTGATGTTCTCTGTTGTTCCCTGGGCTGTGGAAGGAATATCC 420
DB      898 CTTCAGAGTGGGTTTGTGATGTTCTCTGTTGTTCCCTGGGCTGTGGAAGGAATATCC 957
```

```
QY      421 ACTTAGTCCCTACCTGCAATTTCTGAGCTTGTCTTACTCTGTTGCCAATTTGGGCCAAC 480
DB      958 ACTTAGTCCCTACCTGCAATTTCTGAGCTTGTCTTACTCTGTTGCCAATTTGGGCCAAC 1017
QY      481 CGAATTTCTCCCAATCTTTATCTTGGTCCGAGGAGATGTCCTCAACAGAGGCTGATA 540
DB      1018 CGAATTTCTCCCAATCTTTATCTTGGTCCGAGGAGATGTCCTCAACAGAGGCTGATA 1077
QY      541 CAGCAATGGGATGTTATGTTATGTTAAATGCAAGCTATACCTGTCCAAAGCTGACTTT 600
DB      1078 CAGCAATGGGATGTTATGTTATGTTAAATGCAAGCTATACCTGTCCAAAGCTGACTTT 1137
QY      601 ATCCCGAGTCTCATTTCTCGCTGTGCTGGAATGACGTTTGTGAAAAATTTTG 660
DB      1138 ATCCCGAGTCTCATTTCTCGCTGTGCTGGAATGACGTTTGTGAAAAATTTTG 1197
QY      661 CCGTGTGTCACAAATCAGTATGATTTCTTGAAGAAAGAAAGCTCCCAATGATGTGTT 720
DB      1198 CCGTGTGTCACAAATCAGTATGATTTCTTGAAGAAAGAAAGCTCCCAATGATGTGTT 1257
QY      721 CTATGTCATGTTTATGTCGGATCTCCGCTCCGCCAATCGCTATCGCTATCATCATG 780
DB      1258 CTATGTCATGTTTATGTCGGATCTCCGCTCCGCCAATCGCTATCGCTATCATCATG 1317
QY      781 AAGAGATGSAATGCTTTTATGATGAAGCTTAAGATTTGGAAGAAAGAAAGCTTACT 840
DB      1318 AAGAGATGSAATGCTTTTATGATGAAGCTTAAGATTTGGAAGAAAGAAAGCTTACT 1377
QY      841 ATATCTCCAACTTCAATTTTCTGGGCCCACTCTGCACTATGGAAGAAATTAAGAAC 900
DB      1378 ATATCTCCAACTTCAATTTTCTGGGCCCACTCTGCACTATGGAAGAAATTAAGAAC 1437
QY      901 CAGACTGAGATCAGGGGCCAAAGCAAACTCAAGCTGCTGCACTGAGAGCCCAAT 960
DB      1438 CAGACTGAGATCAGGGGCCAAAGCAAACTCAAGCTGCTGCACTGAGAGCCCAAT 1497
QY      961 GAACCTGTCCCTGTGCTGCTCAGAGAGGTGSAAGAGAAAGCGAGAGCCCTCAGTCCACC 1020
DB      1498 GAACCTGTCCCTGTGCTGCTCAGAGAGGTGSAAGAGAAAGCGAGAGCCCTCAGTCCACC 1557
QY      1021 TGTGCCCATCTGTGCTTACCTCAGAGGCAAGCAAAAGCCCGTGCATCCGCCAGCGTG 1080
DB      1558 TGTGCCCATCTGTGCTTACCTCAGAGGCAAGCAAAAGCCCGTGCATCCGCCAGCGTG 1617
QY      1081 CCCAGCGTCCGACCGTGCAGCCCTCGCTGTTAAGAGACGCCGTGTGACAGCGCTC 1140
DB      1618 CCCAGCGTCCGACCGTGCAGCCCTCGCTGTTAAGAGACGCCGTGTGACAGCGCTC 1677
QY      1141 AGTGGGCTGCACCTGTCCGAGACAGGCTGGAAGACAGCAATTAAGTCAAGCGTTCCTTC 1200
DB      1678 AGTGGGCTGCACCTGTCCGAGACAGGCTGGAAGACAGCAATTAAGTCAAGCGTTCCTTC 1737
QY      1201 TCTCTGATATCAAAATCAGTTTCAATTTACGCCAGCATGCGAGATCTTAACTGCTTC 1260
DB      1738 TCTCTGATATCAAAATCAGTTTCAATTTACGCCAGCATGCGAGATCTTAACTGCTTC 1797
QY      1261 TCTCTCATCAAGATGCTTTTGGAAATTAACAACTTCCATCTCTGATGTTGGAACCAAC 1320
DB      1798 TCTCTCATCAAGATGCTTTTGGAAATTAACAACTTCCATCTCTGATGTTGGAACCAAC 1857
QY      1321 AAGCTATGCAATGTTCCCTGTTTACAGGAATTCGAGAGAGACTCCCGGAAACAGTCT 1380
DB      1858 AAGCTATGCAATGTTCCCTGTTTACAGGAATTCGAGAGAGACTCCCGGAAACAGTCT 1917
QY      1381 GATTAAGAGAGAGCCAGCATCCCAAGAGCTGCAAGCCGCAAGCTTTCAGACAGCCAG 1440
DB      1918 GATTAAGAGAGAGCCAGCATCCCAAGAGCTGCAAGCCGCAAGCTTTCAGACAGCCAG 1977
QY      1441 AGCAAGGATGTTGATTCGGTCAAGCAACAGAGAGTGGCAACCGGCCAGAGTCCCTTTTA 1500
DB      1978 AGCAAGGATGTTGATTCGGTCAAGCAACAGAGAGTGGCAACCGGCCAGAGTCCCTTTTA 2037
QY      1501 TCTTCATGCAATGGAAGTGGAGCTGGAGAGCAATTAACACAGCTTCTTTTCGCG 1560
```

Db 2038 TCTCCACTGATGAGTGGAGACGTGGAGACAATTACACACAGCTTCTTTTCGG 2097
QY 1561 CTTTCCACCAAGCCAGACACCTCAGCAAGCTCTGCGCTGGCCTTAAAGGCTGAC 1620
Db 2098 CTTTCCACCAAGCCAGACACCTCAGCAAGCTCTGCGCTGGCCTTAAAGGCTGAC 2157
QY 1621 TCGGATATCTGGCCCCCGAGACCTCAACCCCTTCCCTGACGAGCAGCTGTATTTTGGC 1680
Db 2158 TCGGATATCTGGCCCCCGAGACCTCAACCCCTTCCCTGACGAGCAGCTGTATTTTGGC 2217
QY 1681 ACAGATCTCTCAGACTTCTACTCTGCTCAGCCATCTACGAGAGGAGCTGCACTTACTCT 1740
Db 2218 ACAGATCTCTCAGACTTCTACTCTGCTCAGCCATCTACGAGAGGAGCTGCACTTACTCT 2277
QY 1741 GCTTACAGCTGCAAGCCAGCTGCCCACTTGGGAGACCAAGTATTTCTGTGGCAGGCGG 1800
Db 2278 GCTTACAGCTGCAAGCCAGCTGCCCACTTGGGAGACCAAGTATTTCTGTGGCAGGCGG 2237
QY 1801 CAGAACCCAGTGAAGAGCTGACTGCGCGGCGAGCTGGCATGAAGAGAGCCCTTTTGA 1860
Db 2338 CAGAACCCAGTGAAGAGCTGACTGCGCGGCGAGCTGGCATGAAGAGAGCCCTTTTGA 2397
QY 1861 AACGATTTTAAAGCAGAAAGCTGCCAAATGGAATTTGAGAGAGCATATGTCAGAGAAC 1920
Db 2398 AACGATTTTAAAGCAGAAAGCTGCCAAATGGAATTTGAGAGAGCATATGTCAGAGAAC 2457
QY 1921 AGGTCAAGGAGAGAGCTGGGAAAATGGGAGTCACTTGTCTTTTGGGAGCATGGA 1980
Db 2458 AGGTCAAGGAGAGAGCTGGGAAAATGGGAGTCACTTGTCTTTTGGGAGCATGGA 2517
QY 1981 ATCATTGAGGTCTCC 1995
Db 2518 ATCATTGAGGTCTCC 2532

RESULT 9

US-10-648-593-115
; Sequence 115, Application US/10648593
; Publication No. US20040106132A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION OF GENES FOR PREDICTING ACTIVITY OF COMPOUNDS THAT
; INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE KINASES AND/OR
; TITLE OF INVENTION: PROTEIN TYROSINE KINASE PATHWAYS IN BREAST CELLS
; FILE REFERENCE: D0273 NP
; CURRENT APPLICATION NUMBER: US/10/648,593
; CURRENT FILING DATE: 2003-08-26
; PRIOR APPLICATION NUMBER: 60/406,385
; PRIOR FILING DATE: 2002-08-27
; NUMBER OF SEQ ID NOS: 557
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 115
; LENGTH: 4790
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-648-593-115

Query Match 94.9%; Score 1893; DB 17; Length 4790;

Best Local Similarity 99.9%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCCCATGAGATGATGGAACTCAAAATTTTACTGAGAGTTGTGGCTTGTGGAA 60
Db 184 ATGGCCCATGAGATGATGGAACTCAAAATTTTACTGAGAGTTGTGGCTTGTGGAA 243
QY 61 AGTGGAAAGGAAAGTGTCTTAATGATAGCCGGCAATTTTGGAAATCAATACATCC 120
Db 244 AGTGGAAAGGAAAGTGTCTTAATGATAGCCGGCAATTTTGGAAATCAATACATCC 303
QY 121 CACATTTTGAAGCATTATATCAATGCTCCAGCTTATGAAGCAAGTTGCAACAG 180
Db 304 CACATTTTGAAGCATTATATCAATGCTCCAGCTTATGAAGCAAGTTGCAACAG 363

QY 181 GACAAAGTGTAAATTACAGAGCTCATCCAGCATTCAGCGAAACATTAAGTTGACATTGAT 240
Db 364 GACAAAGTGTAAATTACAGAGCTCATCCAGCATTCAGCGAAACATTAAGTTGACATTGAT 423
QY 241 TGCAGTCAGAAAGTTGTAGTTTACATCAATCAAGCTCCCAAGATGTTGCTCTCTTCA 300
Db 424 TGCAGTCAGAAAGTTGTAGTTTACATCAATCAAGCTCCCAAGATGTTGCTCTCTTCA 483
QY 301 GACTGTTTTCATCTGTACTTCTGGGTAAACTGAGAGAGCTTCAACTCTGTTCACCTG 360
Db 484 GACTGTTTTCATCTGTACTTCTGGGTAAACTGAGAGAGCTTCAACTCTGTTCACCTG 543
QY 361 CTTCAGAGTGGGTTTGTAGTCTCTCGTGTGTTTCCCGGCTCTGTGAAGGAAATCC 420
Db 544 CTTCAGAGTGGGTTTGTAGTCTCTCGTGTGTTTCCCGGCTCTGTGAAGGAAATCC 603
QY 421 ACTCTAGTCCCTACCTGCAATTTCTCAGCCTTCTTACTGTGCAACATTTGGGCAACC 480
Db 604 ACTCTAGTCCCTACCTGCAATTTCTCAGCCTTCTTACTGTGCAACATTTGGGCAACC 663
QY 481 CGAATTTCTCCAAATCTTTATCTTGGCTGCAAGCATGCTCTCAACAGAGCTGATA 540
Db 664 CGAATTTCTCCAAATCTTTATCTTGGCTGCAAGCATGCTCTCAACAGAGCTGATA 723
QY 541 CAGCAAAATGGATTTGTTATGTTAAATGCCAGTATACCTGTCCAAAGCCTGACTT 600
Db 724 CAGCAAAATGGATTTGTTATGTTAAATGCCAGTATACCTGTCCAAAGCCTGACTT 783
QY 601 ATCCCGAGATCATTTCTGCGTGTGCTGTGAATGACAGCTTTTGTGAGAAATTTTG 660
Db 784 ATCCCGAGATCATTTCTGCGTGTGCTGTGAATGACAGCTTTTGTGAGAAATTTTG 843
QY 661 CCGTGGTTGACAAATCAGTGAATTTCAATTGAGAAAGCAAAAGCCTCCAAATGATGTT 720
Db 844 CCGTGGTTGACAAATCAGTGAATTTCAATTGAGAAAGCAAAAGCCTCCAAATGATGTT 903
QY 721 CTAGTGCATGTTTACGTGGAGATCTCCGCTCCGCCACATGCTATGCTTACATCATG 780
Db 904 CTAGTGCATGTTTACGTGGAGATCTCCGCTCCGCCACATGCTATGCTTACATCATG 963
QY 781 AAGAGATGACATGCTTTAGATGAAGCTTACAGATTTGTGAAAGAAAGAAAGCACTACT 840
Db 964 AAGAGATGACATGCTTTAGATGAAGCTTACAGATTTGTGAAAGAAAGAAAGCACTACT 1023
QY 841 ATATCTCCAAACTTCAATTTTCTGGGCCAACTCTGAGCTATGAGAAAGATTAAGAAC 900
Db 1024 ATATCTCCAAACTTCAATTTTCTGGGCCAACTCTGAGCTATGAGAAAGATTAAGAAC 1083
QY 901 CAGACTGAGCATCAGGAGCCAAAGCAAACTCAACTGCTGCACTTGGAGAGAGCCCAAT 960
Db 1084 CAGACTGAGCATCAGGAGCCAAAGCAAACTCAACTGCTGCACTTGGAGAGAGCCCAAT 1143
QY 961 GAACCTGTCCCTGCTCTCAGAGGTTGAGCAGAAAGAGCAGAGCCCTCACTCCACC 1020
Db 1144 GAACCTGTCCCTGCTCTCAGAGGTTGAGCAGAAAGAGCAGAGCCCTCACTCCACC 1203
QY 1021 TGTGCCGACTTGTACTCTCAGAGCAGAGGCAAAAGGCCGTGATCCCGCAGCGTG 1080
Db 1204 TGTGCCGACTTGTACTCTCAGAGCAGAGGCAAAAGGCCGTGATCCCGCAGCGTG 1263
QY 1081 CCCAGGCTGCCAGCGTGCAGCGCTGTGTTTGAAGACAGCCCGCTGTGTAAGGGGCTC 1140
Db 1264 CCCAGGCTGCCAGCGTGCAGCGCTGTGTTTGAAGACAGCCCGCTGTGTAAGGGGCTC 1323
QY 1141 AGTGGCTGCACCTGTCCGAGAGCAGGCTGGAAGCAGCAATTAAGCTCAAGCTTCTTC 1200
Db 1324 AGTGGCTGCACCTGTCCGAGAGCAGGCTGGAAGCAGCAATTAAGCTCAAGCTTCTTC 1383
QY 1201 TCTCTGATATCAAAATCAGTTTCAATTTAGCCAGCATGAGCATCTTACATGCTTC 1260
Db 1384 TCTCTGATATCAAAATCAGTTTCAATTTAGCCAGCATGAGCATCTTACATGCTTC 1443

```
QY 1261 TCCTCATCAGAAAGTGTTCGTAATCTACAAACCTTTCACCTACTCTGTGATGGAGCAAC 1320
    |||||
Db 1444 TCCTCATCAGAAAGTGTTCGTAATCTACAAACCTTTCACCTACTCTGTGATGGAGCAAC 1503
QY 1321 AAGCTATGCCAGTTCTCCCTGTTTCAGAACTATCGAGCAGACTCCGAAACCAAGTCT 1380
    |||||
Db 1504 AACCTATGCCAGTTCTCCCTGTTTCAGAACTATCGAGCAGACTCCGAAACCAAGTCT 1563
QY 1381 GATTAAGGAGGAGCAGCATCCCAAGAACTCCAGACCCGCGCTTCAGACAGCCAG 1440
    |||||
Db 1564 GATTAAGGAGGAGCAGCATCCCAAGAACTCCAGACCCGCGCTTCAGACAGCCAG 1623
QY 1441 AGCAAGGATTTGATTCGATCGATCAGAACCCAGCAGCAGTGGCACCCGCGAGTCTCTTTTA 1500
    |||||
Db 1624 AGCAAGGATTTGATTCGATCGATCAGAACCCAGCAGCAGTGGCACCCGCGAGTCTCTTTTA 1683
QY 1501 TCTTCACCTGCATGAAGTGGAGCGTGGAGGACAAATTACACACAGCTTCTTTTCGCG 1560
    |||||
Db 1684 TCTTCACCTGCATGAAGTGGAGCGTGGAGGACAAATTACACACAGCTTCTTTTCGCG 1743
QY 1561 CTTTCCACCAAGCAGCAGCAGCCTTCAGAAAGTCTGTGCTGCTGGCTTTAAGGCTGGCAC 1620
    |||||
Db 1744 CTTTCCACCAAGCAGCAGCAGCCTTCAGAAAGTCTGTGCTGCTGGCTTTAAGGCTGGCAC 1803
QY 1621 TCGGATATCTTGGCCCCCAGACCTCTACCCCTTCCCTGACACAGCAGCTGTATTTTGGC 1680
    |||||
Db 1804 TCGGATATCTTGGCCCCCAGACCTCTACCCCTTCCCTGACACAGCAGCTGTATTTTGGC 1863
QY 1681 ACAGAGTCTCTACACCTTCTACTCTGCTCAGCAGCAGTCTTACGAGGAGAGTGCAGTTACTCT 1740
    |||||
Db 1864 ACAGAGTCTCTACACCTTCTACTCTGCTCAGCAGCAGTCTTACGAGGAGAGTGCAGTTACTCT 1923
QY 1741 GCTTACAGCTGCGAGCAGCAGCTGCGCCTTCGAGGAGCAAGTCTTCTGTGCGCAGGCGG 1800
    |||||
Db 1924 GCTTACAGCTGCGAGCAGCAGCTGCGCCTTCGAGGAGCAAGTCTTCTGTGCGCAGGCGG 1983
QY 1801 CAGAACCCAGTGAAGCAGAGCTGCTGCGCGGAGCTGCGCATGAAGAGCCCTTTGAA 1860
    |||||
Db 1984 CAGAACCCAGTGAAGCAGAGCTGCTGCGCGGAGCTGCGCATGAAGAGCCCTTTGAA 2043
QY 1861 AAGCAGTTTAAACGAGAAAGCTGCCAAATGGAATTTGAGAGAGCATCTGTCAGAGAAC 1920
    |||||
Db 2044 AAGCAGTTTAAACGAGAAAGCTGCCAAATGGAATTTGAGAGAGCATCTGTCAGAGAAC 2103
QY 1921 AGGTCAAGGGAAGAGCTGGGGAAGGTGGGAGTCACTAGCTTTTCGGGCGAGCATGGA 1980
    |||||
Db 2104 AGGTCAAGGGAAGAGCTGGGGAAGGTGGGAGTCACTAGCTTTTCGGGCGAGCATGGA 2163
QY 1981 ATCATTTAGAGTCTCC 1995
    |||||
Db 2164 ATCATTTAGAGTCTCC 2178

RESULT 10
US-10-296-115-520
; Sequence 520, Application US/10296115
; Publication No. US20040053248A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq Inc
; TITLE OF INVENTION: No. US20040053248A1 Nucleic Acids and Polypeptides
; FILE REFERENCE: 784PCT
; CURRENT APPLICATION NUMBER: US/10/296,115
; PRIOR FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: US09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US09/552,317
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 1478
; SEQ ID NO 520
; LENGTH: 2966
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
```

```
; NAME/KEY: misc.feature
; LOCATION: (1)-(2966)
; OTHER INFORMATION: n = a,t,c or g
US-10-296-115-520

Query Match          92.7%; Score 1849; DB 13; Length 2966;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1949; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 45 GGTGCTCTGCTGGAAGTGGAAACGAAAGTGTCTGTAATGATAGCCGCAATTTGT 104
    |||||
Db 67 GGTGCTCTGCTGGAAGTGGAAACGAAAGTGTCTGTAATGATAGCCGCAATTTGT 126
QY 105 GGAATACATATATATCCCATTTTGGAGGCCATTTATATATCACTGTCCAGCTTATGAA 164
    |||||
Db 127 GGAATACATATATATCCCATTTTGGAGGCCATTTATATATCACTGTCCAGCTTATGAA 186
QY 165 GCGAAGTTGCAACAGACAAAGTGTATTAACAGAGCTCATCCAGCAATTCAGCAAAACA 224
    |||||
Db 187 GCGAAGTTGCAACAGACAAAGTGTATTAACAGAGCTCATCCAGCAATTCAGCAAAACA 246
QY 225 TAAGTTGACATTGATTCAGAGTTCAGAGGTTGATTAAGATCAAAGCTCCAAAGATGT 284
    |||||
Db 247 TAAGTTGACATTGATTCAGAGTTCAGAGGTTGATTAAGATCAAAGCTCCAAAGATGT 306
QY 285 TGGCTCTCTCTTCAGACTGTTTTCTGACTGTACTTCGCGGTAACTGGAGAAAGAGCTT 344
    |||||
Db 307 TGGCTCTCTCTTCAGACTGTTTTCTGACTGTACTTCGCGGTAACTGGAGAAAGAGCTT 366
QY 345 CAACTCTGTTCACCTGTGTCAGAGTGGTTCCTGAGTTCCTCGTTTTCCTGGGCT 404
    |||||
Db 367 CAACTCTGTTCACCTGTGTCAGAGTGGTTCCTGAGTTCCTCGTTTTCCTGGGCT 426
QY 405 CTGTGAAGAAATATCACTCTATGTCCTTACCTGCATTTCTGAGCCTTGTACCTGTTGC 464
    |||||
Db 427 CTGTGAAGAAATATCACTCTATGTCCTTACCTGCATTTCTGAGCCTTGTACCTGTTGC 486
QY 465 CAACAATGGGCAACCCGAAATTTCTCCGAATCTTTATCTGTGCGCCAGCAGAGTCTCT 524
    |||||
Db 487 CAACAATGGGCAACCCGAAATTTCTCCGAATCTTTATCTGTGCGCCAGCAGAGTCTCT 546
QY 525 CAACAAGAGCTGATACAGCAGAAATGGATTTGTTATGTGTTAAATGCACTATATCTG 584
    |||||
Db 547 CAACAAGAGCTGATACAGCAGAAATGGATTTGTTATGTGTTAAATGCACTATATCTG 606
QY 585 TCCAAGCTCTGATTATCCCGAGTCTCATTTCTCGCTGTGCTGTGAATGACAGCTT 644
    |||||
Db 607 TCCAAGCTCTGATTATCCCGAGTCTCATTTCTCGCTGTGCTGTGAATGACAGCTT 666
QY 645 TTGTGAGAAATTTTGGCGTGTGGCAAAATCAATGATGATTTCAATTGAGAAAGCAAAAGC 704
    |||||
Db 667 TTGTGAGAAATTTTGGCGTGTGGCAAAATCAATGATGATTTCAATTGAGAAAGCAAAAGC 726
QY 705 CTCGAATGATGTGTTCTAGTGCACTGTTTATGCTGGATCTCCGCTCCGCAACATGCG 764
    |||||
Db 727 CTCGAATGATGTGTTCTAGTGCACTGTTTATGCTGGATCTCCGCTCCGCAACATGCG 786
QY 765 TATGCTTACATCATGAAGAGATGAGACATGCTTTATGATGAAGCTTACAGATTTGTGAA 824
    |||||
Db 787 TATGCTTACATCATGAAGAGATGAGACATGCTTTATGATGAAGCTTACAGATTTGTGAA 846
QY 825 AGAAAAAGACCTTATATCTCCAAACTTAAATTTTCTGGGCCAATCTCTGAGCTATGA 884
    |||||
Db 847 AGAAAAAGACCTTATATCTCCAAACTTAAATTTTCTGGGCCAATCTCTGAGCTATGA 906
QY 885 GAAAGAGTTAAGAACAGACTGAGCATCAGGCGCAAAAGAGCAAACTCAAGCTGTGCA 944
    |||||
Db 907 GAAAGAGTTAAGAACAGACTGAGCATCAGGCGCAAAAGAGCAAACTCAAGCTGTGCA 966
QY 945 CTGAGAGAGCCAAATGAACTGTCTCTGTCTCAGAGGTTGAGCAAGAAAGCGAGAC 1004
    |||||
Db 967 CTGAGAGAGCCAAATGAACTGTCTCTGTCTCAGAGGTTGAGCAAGAAAGCGAGAC 1026
```


QY 721 CTAGTGCACTGTTAGTGGGATCTCCGCTCCGCCACATCGCTATCGGCTACATCAG 780
DB 1258 CTAGTGCACTGTTAGTGGGATCTCCGCTCCGCCACATCGCTATCGGCTACATCAG 1317
QY 781 AAGAGGATGGAACATGCTTTAGATGAAGACTTAAAGCTTTGAAAGAAAAAGACCTTACT 840
DB 1318 AAGAGGATGGAACATGCTTTAGATGAAGACTTAAAGCTTTGAAAGAAAAAGACCTTACT 1377
QY 841 ATATCTCCAAACTTCAATTTCTGGGCCACTCTGGAATATGAAGAAATTAAGAAC 900
DB 1378 ATATCTCCAAACTTCAATTTCTGGGCCACTCTGGAATATGAAGAAATTAAGAAC 1437
QY 901 CAGACTGAGCATCAGGGCCAAAGAGCAATCAAGCTGTGACCTGAGAGCAACCAAT 960
DB 1438 CAGACTGAGCATCAGGGCCAAAGAGCAATCAAGCTGTGACCTGAGAGCAACCAAT 1497
QY 961 GAACTGTCTCTGTCTCAGAGGGTGAACAGAAAAAGCGAGACGCTCTCACTCCACC 1020
DB 1498 GAACTGTCTCTGTCTCAGAGGGTGAACAGAAAAAGCGAGACGCTCTCACTCCACC 1557
QY 1021 TGTGCCGACT 1080
DB 1558 TGTGCCGACT 1617
QY 1081 CCCAGGCTGCGCCAGCGCTGACGCGCTGTGAGAGACAGCCGCTGTGAGAGAGGCTC 1140
DB 1618 CCCAGGCTGCGCCAGCGCTGACGCGCTGTGAGAGACAGCCGCTGTGAGAGAGGCTC 1677
QY 1141 AGTGGCTGTGACCTGTCTCCGAGACAGGCTGGAAGACAGCAATAGCTCAAGCGTTCTTC 1200
DB 1678 AGTGGCTGTGACCTGTCTCCGAGACAGGCTGGAAGACAGCAATAGCTCAAGCGTTCTTC 1737
QY 1201 TCTCTGATATCAAACTCAATTCATATTCAGCCAGATGAGCAATCTCTTAATGCTTTC 1260
DB 1738 TCTCTGATATCAAACTCAATTCATATTCAGCCAGATGAGCAATCTCTTAATGCTTTC 1797
QY 1261 TCTCTGATATCAAACTCAATTCATATTCAGCCAGATGAGCAATCTCTTAATGCTTTC 1320
DB 1798 TCTCTGATATCAAACTCAATTCATATTCAGCCAGATGAGCAATCTCTTAATGCTTTC 1857
QY 1321 AAGCTATGCGAGTTCTCCCTGTTTCAAGAACTATCGAGACAGCTCCGAAACCACTCTCT 1380
DB 1858 AAGCTATGCGAGTTCTCCCTGTTTCAAGAACTATCGAGACAGCTCCGAAACCACTCTCT 1917
QY 1381 GATTAAGAGAGAGGCGAGCATCTCCCAAGAGCTGACACCGGCTTCAAGACAGCGAG 1440
DB 1918 GATTAAGAGAGAGGCGAGCATCTCCCAAGAGCTGACACCGGCTTCAAGACAGCGAG 1977
QY 1441 AGCAAGCATTTGCAATTCGGTCAAGAACAGAGAGAGTGGACCGCCAGAGGTCCTTTTA 1500
DB 1978 AGCAAGCATTTGCAATTCGGTCAAGAACAGAGAGAGTGGACCGCCAGAGGTCCTTTTA 2037
QY 1501 TCTCCACTGATCGAAGTGGAGCGTGAAGACAAATTACACACAGCTTCTTTTCGGC 1560
DB 2038 TCTCCACTGATCGAAGTGGAGCGTGAAGACAAATTACACACAGCTTCTTTTCGGC 2097
QY 1561 CTTTTCACAGCAGAGACAGCACTCAAGAGTCTGCTGGGCTTTAAGGGCTGGCAC 1620
DB 2098 CTTTTCACAGCAGAGACAGCACTCAAGAGTCTGCTGGGCTTTAAGGGCTGGCAC 2157
QY 1621 TCGGATATCTTGGCCCCCGACAGCTTACCCCTTCCCTGACAGCAGCTGATTTTGGC 1680
DB 2158 TCGGATATCTTGGCCCCCGACAGCTTACCCCTTCCCTGACAGCAGCTGATTTTGGC 2217
QY 1681 ACAGAGTCTCTCACTTCTTACTCTGCTGAGCATCTACGAGAGAGTGGCCAGTTACTCT 1740
DB 2218 ACAGAGTCTCTCACTTCTTACTCTGCTGAGCATCTACGAGAGAGTGGCCAGTTACTCT 2277
QY 1741 GGTCTACAGTGAAGCAGAGTGGCCAGTTTGGGAGACCAAGTCTATTTCTGCGCAGGCG 1800
DB 2278 GGTCTACAGTGAAGCAGAGTGGCCAGTTTGGGAGACCAAGTCTATTTCTGCGCAGGCG 2337

QY 1801 CAGAAAGCAGAGTGAAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1860
DB 2338 CAGAAAGCAGAGTGAAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 2397
QY 1861 AAGCAGTTTAAACGAGAGAGTGGCCAAATGGAATTTGAGAGAGATCATGTCAGAGAAC 1920
DB 2398 AAGCAGTTTAAACGAGAGAGTGGCCAAATGGAATTTGAGAGAGATCATGTCAGAGAAC 2457
QY 1921 AGGTACGGGAAAGAGCTGGGAAAGTGGGAGTCACTAGCTTTTCGGGAGCATGGAA 1980
DB 2458 AGGTACGGGAAAGAGCTGGGAAAGTGGGAGTCACTAGCTTTTCGGGAGCATGGAA 2517
QY 1981 ATCATTGAGTCTCC 1995
DB 2518 ATCATTGAGTCTCC 2532

RESULT 12
US-10-094-749-673
; Sequence 673, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TARAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOMYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 08435/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 673
; LENGTH: 2102
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-673

Query Match 89.8%; Score 1791; DB 16; Length 2102;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1991; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAGGCCCATGAGATGATGGAACCAATTGTAAGAGAGTGGTGGCTCTGCTGGA 60
DB 56 AAGGCCCATGAGATGATGGAACCAATTGTAAGAGAGTGGTGGCTCTGCTGGA 115
QY 61 AAGTGAACGAAAAAGTGTGCTAATGATAGCCGCAATTTGGAATACAAATACATCC 120
DB 116 AAGTGAACGAAAAAGTGTGCTAATGATAGCCGCAATTTGGAATACAAATACATCC 175
QY 121 CACATTTTGAAGACCATTAATATCACTGCTCAAGCTTATGAAGCGAAGTTGCAACAG 180
DB 176 CACATTTTGAAGACCATTAATATCACTGCTCAAGCTTATGAAGCGAAGTTGCAACAG 235
QY 181 GACAAAGTGTAAATTAACAGAGCTCATCGAGCAATTCAGGAAACATAAGGTTGACATTGAT 240

Db 236 GACAAAGTGTATTAACAGAGCTCATCCAGCAATTCAGCGAAACATTAAGGTGACATTTGAT 295
QY 241 TGCAGTCAGAAAGTTGATGATTAAGATCAAAAGTCCCAAGATTTGCTCTCTCTCA 300
Db 296 TGCAGTCAGAAAGTTGATGATTAAGATCAAAAGTCCCAAGATTTGCTCTCTCTCA 355
QY 301 GACTGTTTCTCACTGTAATCTTGGGTAAACAGGAAAGAGCTCAACTCTGTTCACTG 360
Db 356 GATGTTTCTCACTGTAATCTTGGGTAAACAGGAAAGAGCTCAACTCTGTTCACTG 415
QY 361 CTTCAGAGTGGGTTTGGTGAATCTCTGTTGTTTCCTGGCCCTCTGTGAAGGAAATCC 420
Db 416 CTTCAGAGTGGGTTTGGTGAATCTCTGTTGTTTCCTGGCCCTCTGTGAAGGAAATCC 475
QY 421 ACTCTAGTCCCTACCTGCAATTTCTCAGCTCTTACCTGTTGCCAAATGGGGCCAAAC 480
Db 476 ACTCTAGTCCCTACCTGCAATTTCTCAGCTCTTACCTGTTGCCAAATGGGGCCAAAC 535
QY 481 CGAATTCCTCCCAATGTTTATCTTGGCTGCGAGCGAGATGCTCAACAAAGAGCTGATA 540
Db 536 CTATATCTTCCCAATCTTATCTTGGCTGCGAGCGAGATGCTCAACAAAGAGCTGATA 595
QY 541 CAGCAAAATGGGATTTGGTATGTTAAATGCGAGTATACCTGTCCAAAGCCTGACTTT 600
Db 596 CAGCAAAATGGGATTTGGTATGTTAAATGCGAGTATACCTGTCCAAAGCCTGACTTT 655
QY 601 ATCCCGAGTCTCATTTCTGCGCTGCTGCTGATGATGACAGCTTTTGTGAAATTTTG 660
Db 656 ATCCCGAGTCTCATTTCTGCGCTGCTGCTGATGATGACAGCTTTTGTGAAATTTTG 715
QY 661 CCGTGGTTGGAACAATCAGTAGATTTTCAATGAAAGAAAGAAACCTCCAAATGATGTT 720
Db 716 CCGTGGTTGGAACAATCAGTAGATTTTCAATGAAAGAAAGAAACCTCCAAATGATGTT 775
QY 721 CTAGTCACTGTTTAACTGGGATCTCCCGCTCCGCCACCATCGCTATGCGCTACATCATG 780
Db 776 CTAGTCACTGTTTAACTGGGATCTCCCGCTCCGCCACCATCGCTATGCGCTACATCATG 835
QY 781 AAGAGATGACATGTTTAAATGATGAAGCTTACAGATTTGTGAAAGAAAGAAAGAAAG 840
Db 836 AAGAGATGACATGTTTAAATGATGAAGCTTACAGATTTGTGAAAGAAAGAAAGAAAG 895
QY 841 ATATCTCCAAATTTCAATTTTCTGGGCAACTCTGATCTATGAAAGAAAGATTAAGAAC 900
Db 896 ATATCTCCAAATTTCAATTTTCTGGGCAACTCTGATCTATGAAAGAAAGATTAAGAAC 955
QY 901 CAGACTGAGCATCAGGGCCAAAGAGAAATCTCAAGCTGCTGCACTGGAAGAGCAAAAT 960
Db 956 CAGACTGAGCATCAGGGCCAAAGAGAAATCTCAAGCTGCTGCACTGGAAGAGCAAAAT 1015
QY 961 GAACTGTCCCTGCTCTCTCAAGGGTGAACAGAAAGCGAGACGCCCTCACTCCACCC 1020
Db 1016 GAACTGTCCCTGCTCTCTCAAGGGTGAACAGAAAGCGAGACGCCCTCACTCCACCC 1075
QY 1021 TGTGCCGACTCTCTCTCTCTCAAGGGTGAACAGAAAGCGCTGATCTCCGCAAGCGT 1080
Db 1076 TGTGCCGACTCTCTCTCTCTCTCAAGGGTGAACAGAAAGCGCTGATCTCCGCAAGCGT 1135
QY 1081 CCCAGGCTGCGAGCGTGAACCGCTGCTTGAAGAGACAGCCGCTGGTGAACAGGGGCTC 1140
Db 1136 CCCAGGCTGCGAGCGTGAACCGCTGCTTGAAGAGACAGCCGCTGGTGAACAGGGGCTC 1195
QY 1141 AGTGGGCTGCACTGTCTCCGAGACAGAGCTGGAAGACAGCATTAAGCTCAAGCTTCCTC 1200
Db 1196 AGTGGGCTGCACTGTCTCCGAGACAGAGCTGGAAGACAGCATTAAGCTCAAGCTTCCTC 1255
QY 1201 TCTCTGATATCAAAATCAATTTTCAATTTTCAAGCGACATGACATCTTCAATGCTTC 1260
Db 1256 TCTCTGATATCAAAATCAATTTTCAATTTTCAAGCGACATGACATCTTCAATGCTTC 1315
QY 1261 TCTCTATCAGAAAGTGTGAAATCTAACAACCTTCACTCTGATGGAGCAAC 1320

Db 1316 TCTCTATCAGAAAGTGTGAAATCTAACAACCTTCACTCTGATGGAGCAAC 1375
QY 1321 AAGCTATGCAATTTCTCCCTGTTTCAAGAACTATCGAGACACATCCCGAAACAGTCC 1380
Db 1376 AAGCTATGCAATTTCTCCCTGTTTCAAGAACTATCGAGACACATCCCGAAACAGTCC 1435
QY 1381 GATTAAGAGAAAGCCAGCATCTCCCAAGAGCTGACAGCCGAGGCTTCAAGACACAG 1440
Db 1436 GATTAAGAGAAAGCCAGCATCTCCCAAGAGCTGACAGCTGACAGGCTTCAAGACACAG 1495
QY 1441 AGCAAGCATTCATTTGGTGAAGAACAGAGACAGTGGGACCGCCAGAGTCCCTTTTA 1500
Db 1496 AGCAAGCATTCATTTGGTGAAGAACAGAGACAGTGGGACCGCCAGAGTCCCTTTTA 1555
QY 1501 TCTCCACTGCAATCGAAGTGGAGCGGTGAGAACATTAACACACAGCTTCTTTTCCGC 1560
Db 1556 TCTCCACTGCAATCGAAGTGGAGCGGTGAGAACATTAACACACAGCTTCTTTTCCGC 1615
QY 1561 CTTTCCACGAGCAGACGACCTCAAGAGTGTGCTGGCTGGGCTTAAAGGCTGGCAC 1620
Db 1616 CTTTCCACGAGCAGACGACCTCAAGAGTGTGCTGGCTGGGCTTAAAGGCTGGCAC 1675
QY 1621 TCGAATATCTTGAGCCGCCAGACCTCTACCCCTTCCCTGACAGACAGCTGTATTGGC 1680
Db 1676 TCGAATATCTTGAGCCGCCAGACCTCTACCCCTTCCCTGACAGACAGCTGTATTGGC 1735
QY 1681 ACAGATCTTCACTTCTACTGTGCTGACGCACTTACGAGAGAGAGTGGCCAGTTACTCT 1740
Db 1736 ACAGATCTTCACTTCTACTGTGCTGACGCACTTACGAGAGAGAGTGGCCAGTTACTCT 1795
QY 1741 GCTACAGCTGAGCAGCTGCGCACTTGGAGAGACCAAGTCTATTCTGTGGCAGGCGG 1800
Db 1796 GCTACAGCTGAGCAGCTGCGCACTTGGAGAGACCAAGTCTATTCTGTGGCAGGCGG 1855
QY 1801 CAGAAAGCAATGACAGAGCTGACTCGCGCGAGCTGACATGAAGAGACCCCTTTGAA 1860
Db 1856 CAGAAAGCAATGACAGAGCTGACTCGCGCGAGCTGACATGAAGAGAGCCCTTTGAA 1915
QY 1861 AAGCATTTAAACGAGAAAGCTGCCAAATGGAATTTGGAGAGACATCATGTCAGAGAAC 1920
Db 1916 AAGCATTTAAACGAGAAAGCTGCCAAATGGAATTTGGAGAGACATCATGTCAGAGAAC 1975
QY 1921 AGGTACCGGGAAGAGCTGGGGAAGGTGGGCACTGCTAGCTTTTGGGAGACATGGA 1980
Db 1976 AGGTACCGGGAAGAGCTGGGGAAGGTGGGCACTGCTAGCTTTTGGGAGACATGGA 2035
QY 1981 ATCATTTAGAGTCTCC 1995
Db 2036 ATCATTTAGAGTCTCC 2050

RESULT 13
US-09-964-277-20
; Sequence 20, Application us/09964277
; Patent No. US20020137170A1
; GENERAL INFORMATION:
; APPLICANT: Luche, Ralf M.
; TITLE OF INVENTION: DSP-16 DUAL-SPECIFICITY PHOSPHATASE
; FILE REFERENCE: 200125_434
; CURRENT APPLICATION NUMBER: US/09/964,277
; CURRENT FILING DATE: 2001-09-25
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 3332
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-964-277-20

Query_Match 71.1%; Score 1418; DB 9; Length 3332;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1418; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 578 ATACCTGTCCAAAGCCGTAATTATCCCCGAGCTCAATTTCTGCGGTGCTGTGAATG 637
DB 975 ATATCTGTCCAAAGCCGTAATTATCCCCGAGCTCAATTTCTGCGGTGCTGTGAATG 1034
QY 638 ACAGCTTTTGTGAAAAATTTTGCCTGTTGGAACAAATCAGTAGATTTCAATTGAGAAG 697
DB 1035 ACAGCTTTTGTGAAAAATTTTGCCTGTTGGAACAAATCAGTAGATTTCAATTGAGAAG 1094
QY 698 CAAAAGCTTCCAAATGATGTTGTTCTAGTGCACTGTTTAACTGCGATCTCCGCTCCGCA 757
DB 1095 CAAAAGCTTCCAAATGATGTTGTTCTAGTGCACTGTTTAACTGCGATCTCCGCTCCGCA 1154
QY 758 CCATCGCTATCGCTCATCATGAAAGATGAGATGAGCAATGCTTTAATGAACTTAAGAT 817
DB 1155 CCATCGCTATCGCTCATCATGAAAGATGAGATGAGCAATGCTTTAATGAACTTAAGAT 1214
QY 818 TTGTGAAAGAAAAAGCCTACTATATCTCCAACTTCAATTTTCTGCGCCAACTCCTG 877
DB 1215 TTGTGAAAGAAAAAGCCTACTATATCTCCAACTTCAATTTTCTGCGCCAACTCCTG 1274
QY 878 ACTATGAGAAAGATTAGAACCAAGCTGAGCATGAGGCGCAAGCAAACTCAAGC 937
DB 1275 ACTATGAGAAAGATTAGAACCAAGCTGAGCATGAGGCGCAAGCAAACTCAAGC 1334
QY 938 TGCTGACCTGAGAAAGCCAAATGAACCTGTCCGTGCTGTCTGAGAGGGTGGACAGAAA 997
DB 1335 TGCTGACCTGAGAAAGCCAAATGAACCTGTCCGTGCTGTCTGAGAGGGTGGACAGAAA 1394
QY 998 GCGAGAAGCCCTCAGTCCACCTGTGCGCACTCTGCTACTCTGAGAGGAGAGAGCAAA 1057
DB 1395 GCGAGAAGCCCTCAGTCCACCTGTGCGCACTCTGCTACTCTGAGAGGAGAGAGCAAA 1454
QY 1058 GGGCCGTGATCCCCGAGCGTSCCAAGCGTGCCTGAGCGTGCAGCCGTGCTGTTAGAG 1117
DB 1455 GGGCCGTGATCCCCGAGCGTSCCAAGCGTGCCTGAGCGTGCAGCCGTGCTGTTAGAG 1514
QY 1118 ACAGCCCGCTGATGAGGGGCTCAGTGGGCTGCACTGTCCGAGAGAGGCTGGAAGAA 1177
DB 1515 ACAGCCCGCTGATGAGGGGCTCAGTGGGCTGCACTGTCCGAGAGAGGCTGGAAGAA 1574
QY 1178 GCAATAGCTCAAGCGTCTCTCTCTGATATCAAAATCAGTTTCATTTAGCCAGCA 1237
DB 1575 GCAATAGCTCAAGCGTCTCTCTCTGATATCAAAATCAGTTTCATTTAGCCAGCA 1634
QY 1238 TGGCAGCATCTTCACTGCTTCTCTCTCATCAGAAATGCTTTGGAATCTAACAACCTT 1297
DB 1635 TGGCAGCATCTTCACTGCTTCTCTCTCATCAGAAATGCTTTGGAATCTAACAACCTT 1694
QY 1298 CCACTACTCTGATGAGGACCAACAGCTATGCAAGTCTCCCTGTTTCAAGAACTATCGG 1357
DB 1695 CCACTACTCTGATGAGGACCAACAGCTATGCAAGTCTCCCTGTTTCAAGAACTATCGG 1754
QY 1358 AGCAGACTCCGAAACAGTCTCTGATAGAGAGAAAGCCAGCATCCCAAGAACTGCGA 1417
DB 1755 AGCAGACTCCGAAACAGTCTCTGATAGAGAGAAAGCCAGCATCCCAAGAACTGCGA 1814
QY 1418 CCGCCAGGCTTCAAGCAGCCAGAGAACGATTTGATTGCTGCTGAGAACCAAGACAGTG 1477
DB 1815 CCGCCAGGCTTCAAGCAGCCAGAGAACGATTTGATTGCTGCTGAGAACCAAGACAGTG 1874
QY 1478 GCACGCGCCAGAGGTCCTTTTATCTCCAGTGCATGAAATGAGGAGCGTGGAGCAATT 1537
DB 1875 GCACGCGCCAGAGGTCCTTTTATCTCCAGTGCATGAAATGAGGAGCGTGGAGCAATT 1934
QY 1538 ACCACACAGCTTCTTTTCCGCTTTCACAGCAGCAGAGCACTCAAGAGTCTGCTG 1597
DB 1935 ACCACACAGCTTCTTTTCCGCTTTCACAGCAGCAGAGCACTCAAGAGTCTGCTG 1994
QY 1598 GCTTGGGCTTAAAGGCTTGGCACTGGATATCTTTGGCCCCCAGACCTTCACTTCC 1657
DB 1995 GCTTGGGCTTAAAGGCTTGGCACTGGATATCTTTGGCCCCCAGACCTTCACTTCC 2054

QY 1658 TGACCAAGCAGCTGATTTTGGCCAGAGATCTCTCACTTTACTTGTGCTAGCCATCT 1717
DB 2055 TGACCAAGCAGCTGATTTTGGCCAGAGATCTCTCACTTTACTTGTGCTAGCCATCT 2114
QY 1718 ACAGAGCAGTSCCAGTTACTCTGCTTACAGCTGAGCAGCTGCGCACTTGGAGACC 1777
DB 2115 ACAGAGCAGTSCCAGTTACTCTGCTTACAGCTGAGCAGCTGCGCACTTGGAGACC 2174
QY 1778 AAGTATTTCTGTGCGCAGGCGCGCAAGCAAGTGAAGAGCTGACTTCCGCGAGCT 1837
DB 2175 AAGTATTTCTGTGCGCAGGCGCGCAAGCAAGTGAAGAGCTGACTTCCGCGAGCT 2234
QY 1838 GGCATGAGAGAGCCCTTTGAAAGCAGTTTAAACGAGAGCTGCCAAATGGAATTG 1897
DB 2235 GGCATGAGAGAGCCCTTTGAAAGCAGTTTAAACGAGAGCTGCCAAATGGAATTG 2294
QY 1898 GAGAGAGCATCTCTCAGAGAACAGTCAAGGAGAGAGCTGGGAAATGGGAGTCACT 1957
DB 2295 GAGAGAGCATCTCTCAGAGAACAGTCAAGGAGAGAGCTGGGAAATGGGAGTCACT 2354
QY 1958 CTAGCTTTTCCGCGCAGCATGGAATCATTTAGGTCTCC 1995
DB 2355 CTAGCTTTTCCGCGCAGCATGGAATCATTTAGGTCTCC 2392
RESULT 14
US-10-072-012-257
Sequence 257, Application US/10072012
GENERAL INFORMATION:
APPLICANT: Tcherny, Velizar
APPLICANT: Zernusen, Bryan
APPLICANT: Patuturajan, Meera
APPLICANT: Shinkets, Richard
APPLICANT: Li, Li
APPLICANT: Gangolli, Esha
APPLICANT: Padigaru, Muralidhara
APPLICANT: Anderson, David W.
APPLICANT: Rastelli, Luca
APPLICANT: Miller, Charles E.
APPLICANT: Gerlach, Valerie
APPLICANT: Taubier Jr, Raymond J.
APPLICANT: Gusev, Vladimir Y.
APPLICANT: Coleman, Steven D.
APPLICANT: Wolenc, Adam R.
APPLICANT: Pena, Carol E. A
APPLICANT: Furtak, Katarzyna
APPLICANT: Grosse, William M.
APPLICANT: Alsebrook II, John P.
APPLICANT: Lepley, Denise M.
APPLICANT: Rieger, Daniel K.
APPLICANT: Burgess, Catherine E.
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-258
CURRENT APPLICATION NUMBER: US/10/072,012
CURRENT FILING DATE: 2002-01-31
PRIOR APPLICATION NUMBER: 60/265,102
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: 60/265,514
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,517
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,412
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,395
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/266,406
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 60/266,767
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: 60/267,057
PRIOR FILING DATE: 2001-02-07

PRIOR APPLICATION NUMBER: 60/266,975
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/267,459
PRIOR FILING DATE: 2001-02-08
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 1391
SOFTWARE: Patent Ver. 2.1
SEQ ID NO 257
LENGTH: 2071
TYPE: DNA
ORGANISM: Homo sapiens
US-10-072-012-257

Query Match 61.8%; Score 1233; DB 13; Length 2071;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1983; Conservative 0; Mismatches 3; Indels 9; Gaps 1;

QY 1 ATGGCCCATGAGATGATGAGAACTCAAAATTGTTACTGAGAGGTGGTGGCTGCTGAGAA 60
DB 61 ATGGCCCATGAGATGATGAGAACTCAAAATTGTTACTGAGAGGTGGTGGCTGCTGAGAA 120
QY 61 AGTGAACGGAAGAGTGTCTGATTAATGATAGCCGCAATTTGTGAAATACATACC 120
DB 121 AGTGAACGGAAGAGTGTCTGATTAATGATAGCCGCAATTTGTGAAATACATACC 180
QY 121 CACATTTTGAAGCCATTAATATCACTGCTCCAACTTTAGAAAGCAAGTTGCAACAG 180
DB 181 CACATTTTGAAGCCATTAATATCACTGCTCCAACTTTAGAAAGCAAGTTGCAACAG 240
QY 181 GACAAAGTGTAAATACAGAGCTCATCAGCAATTCAGCAACATTAAGTTGACATGAT 240
DB 241 GACAAAGTGTAAATACAGAGCTCATCAGCAATTCAGCAACATTAAGTTGACATGAT 300
QY 241 TGCAGTCAGAGGTGTGATTTACGATCAAGCTCCCAAGATGTGGCTCTCTCTTCA 300
DB 301 TGCAGTCAGAGGTGTGATTTACGATCAAGCTCCCAAGATGTGGCTCTCTCTTCA 360
QY 301 GACTGTTTCTCACTGACTCTTGGGTTAACTGGAAGAGCTTCACTCTGTTCACTG 360
DB 361 GACTGTTTCTCACTGACTCTTGGGTTAACTGGAAGAGCTTCACTCTGTTCACTG 420
QY 361 CTTCAGAGTGGGTTGCTGAGTCTCTGATTTTCCCTGGCCCTGTGAAAGAAATCC 420
DB 421 CTTCAGAGTGGGTTGCTGAGTCTCTGATTTTCCCTGGCCCTGTGAAAGAAATCC 480
QY 421 ACTCTAGTCCCTACCTGCAATTTCTCAGCCTTGTGCTTACCTGTGCAACATTTGGCCAC 480
DB 481 ACTCTAGTCCCTACCTGCAATTTCTCAGCCTTGTGCTTACCTGTGCAACATTTGGCCAC 540
QY 481 CGAATTTCTCCAAATCTTAACTTTGGCTGCCAGGAGATGCTCTCAACAGAGCTGATG 540
DB 541 CGAATTTCTCCAAATCTTAACTTTGGCTGCCAGGAGATGCTCTCAACAGAGCTGATG 600
QY 541 CACAGAAATGGAGTGGTTATGTTTAAATGSCAGCTATACCTGTCCAAAGCCGACTT 600
DB 601 CACAGAAATGGAGTGGTTATGTTTAAATGSCAGCTATACCTGTCCAAAGCCGACTT 660
QY 601 ATCCCGAGTCTCATTTCTCGCGTGTGCTGTGAATACAGCTTTGTGAAATTTTG 660
DB 661 ATCCCGAGTCTCATTTCTCGCGTGTGCTGTGAATACAGCTTTGTGAAATTTTG 720
QY 661 CCGTGTGGACAAATCAGTAAATTTCAATGAGAAACAAAGCCCTCAATGATGTT 720
DB 721 CCGTGTGGACAAATCAGTAAATTTCAATGAGAAACAAAGCCCTCAATGATGTT 780
QY 721 CTAGTGAAGTGTAGTGGGATCTCCGCTCCGSCCAATGCTATGCTATCATGATG 780
DB 781 CTAGTGAAGTGTAGTGGGATCTCCGCTCCGSCCAATGCTATGCTATCATGATG 840
QY 781 AAGAGATGACATGCTTTAGATGAAGCTTACAGATTTGTGAAAGAAAGACCTACT 840
DB 841 AAGAGATGACATGCTTTAGATGAAGCTTACAGATTTGTGAAAGAAAGACCTACT 900

QY 841 AATCTCCAAATCTCAATTTTCTGGSCAACTCTGAGCTATGAGAAAGATTAGAAC 900
DB 901 AATCTCCAAATCTCAATTTTCTGGSCAACTCTGAGCTATGAGAAAGATTAGAAC 960
QY 901 CAGACTGAGATCAAGGCGCAAGCAAACTCAAGCTCTGCACTGAGAAAGCAAT 960
DB 961 CAGACTGAGATCAAGGCGCAAGCAAACTCAAGCTCTGCACTGAGAAAGCAAT 1020
QY 961 GAACTGTCTCTGCTCTCAAGAGGTGACAGAAAGGAGAGAGCCCTCACTCAACC 1020
DB 1021 GAACTGTCTCTGCTCTCAAGAGGTGACAGAAAGGAGAGAGCCCTCACTCAACC 1080
QY 1021 TGTGCGGACCTGCTGCTCAAGAGGTGACAGAAAGGAGAGAGCCCTCACTCAACC 1080
DB 1081 TGTGCGGACCTGCTGCTCAAGAGGTGACAGAAAGGAGAGAGCCCTCACTCAACC 1132
QY 1081 CCCAGCGTCCCAAGCTGAGAGCTGCTGCTTAAAGAGAGAGCCCTGCTGATCAGCGCTC 1140
DB 1133 -CCAGCGTCCCAAGCTGAGAGCTGCTGCTTAAAGAGAGAGCCCTGCTGATCAGCGCTC 1191
QY 1141 AGTGGCTGCACTGTCCGCAAGAGCTGAGAAAGCAATTAAGCTCAAGCTTCTTC 1200
DB 1192 AGTGGCTGCACTGTCCGCAAGAGCTGAGAAAGCAATTAAGCTCAAGCTTCTTC 1251
QY 1201 TCTCTGATATCAAAATCAGTTTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1260
DB 1252 TCTCTGATATCAAAATCAGTTTCAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1311
QY 1261 TCTCTATCAGAAAGATCTTGTGAATCTCAAACTTCTCACTCTGATGAGAGAGAG 1320
DB 1312 TCTCTATCAGAAAGATCTTGTGAATCTCAAACTTCTCACTCTGATGAGAGAGAG 1371
QY 1321 AAGCTATGCAAGTCTTCCCTGTTCAAGAACTATGAGAGAGAGAGAGAGAGAGAG 1380
DB 1372 AAGCTATGCAAGTCTTCCCTGTTCAAGAACTATGAGAGAGAGAGAGAGAGAGAG 1431
QY 1381 GATTAAG 1440
DB 1432 GATTAAG 1491
QY 1441 AGCAAGCATTTGATTCGCTCAAGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
DB 1492 AGCAAGCATTTGATTCGCTCAAGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1551
QY 1501 TCTTCACTGATCAAG 1560
DB 1552 TCTTCACTGATCAAG 1611
QY 1561 CTTTCAAG 1620
DB 1612 CTTTCAAG 1671
QY 1621 TCGGATATCTTGGCCCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1680
DB 1672 TCGGATATCTTGGCCCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1731
QY 1681 ACAGAGTCTCACTTCTACTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
DB 1732 ACAGAGTCTCACTTCTACTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1791
QY 1741 GCTTACAGTCAAG 1800
DB 1792 GCTTACAGTCAAG 1851
QY 1801 CAGAGCCAAAGTGAAG 1860
DB 1852 CAGAGCCAAAGTGAAG 1911
QY 1861 AAGCAATTTAAAGCAAG 1920
DB 1912 AAGCAATTTAAAGCAAG 1971
QY 1921 AGGTACGAG 1980

Db 1972 AGGTCACGGGAGAGCTGGGGAAGTGGCAGTCACTAGCTTTTGGGCGAGATGAA 2031
QY 1981 ATCATTTAGGCTCTCC 1995
Db 2032 ATCATTTAGGCTCTCC 2046

RESULT 15

US-10-072-012-255

Sequence 255, Application US/10072012

Publication No. US20040033493A1

GENERAL INFORMATION:

APPLICANT: Tchermeyev, Velizar
APPLICANT: Spytek, Kimberly
APPLICANT: Zerhusen, Bryan
APPLICANT: Patturajan, Meera
APPLICANT: Shinkets, Richard
APPLICANT: Li, Li
APPLICANT: Gangoli, Esha
APPLICANT: Padigaru, Muralidhara
APPLICANT: Anderson, David W.
APPLICANT: Rastelli, Luca
APPLICANT: Miller, Charles E.
APPLICANT: Gerlach, Valerie
APPLICANT: Taupier Jr, Raymond J.
APPLICANT: Gusev, Vladimir Y.
APPLICANT: Colman, Steven D.
APPLICANT: Wolenc, Adam R.
APPLICANT: Pena, Carol E. A
APPLICANT: Furtak, Katarzyna
APPLICANT: Grosse, William M.
APPLICANT: Alsobrook II, John P.
APPLICANT: Lepley, Denise M.
APPLICANT: Rieger, Daniel K.
APPLICANT: Burgess, Catherine E.
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-258
CURRENT APPLICATION NUMBER: US/10/072,012
CURRENT FILING DATE: 2002-01-31
PRIOR APPLICATION NUMBER: 60/265,102
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: 60/265,514
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,517
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,412
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,395
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/266,406
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: 60/266,767
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: 60/267,057
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/266,975
PRIOR FILING DATE: 2001-02-07
PRIOR APPLICATION NUMBER: 60/267,459
PRIOR FILING DATE: 2001-02-08
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 1391
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 255
LENGTH: 2200
TYPE: DNA
ORGANISM: Homo sapiens
US-10-072-012-255

Query Match 59.2%; Score 1182; DB 13; Length 2200;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1182; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 814 AGATTTGTGAAGAAAAAGACCTAATAATCTCCAACTTCAATTTTCTGGGCCAACTC 873
Db 919 AGATTTGTGAAGAAAAAGACCTAATAATCTCCAACTTCAATTTTCTGGGCCAACTC 978
QY 874 CTGACCTATGAGAAAGATTAAAGACAGACTGAGATCAGGGCCAAAGAGCAAACTC 933
Db 979 CTGACCTATGAGAAAGATTAAAGACAGACTGAGATCAGGGCCAAAGAGCAAACTC 1038
QY 934 AAGCTGCTGCACCTGGAAGAGCCAAATGAACCTGCTCCTGCTGTCTCAGAGGGTGAACG 993
Db 1039 AAGCTGCTGCACCTGGAAGAGCCAAATGAACCTGCTCCTGCTGTCTCAGAGGGTGAACG 1098
QY 994 AAAAGCGAGACGCCCTCACTCACTCACTGAGCCAGCTCTGCTCACTCACTCACTCACTCA 1053
Db 1099 AAAAGCGAGACGCCCTCACTCACTCACTGAGCCAGCTCTGCTCACTCACTCACTCACTCA 1158
QY 1054 CAAAGGCCGTGTCATCCCGCAGCGTGCAGCGTGCAGCGTGCAGCGTGCAGCGTGCAGCGT 1113
Db 1159 CAAAGGCCGTGTCATCCCGCAGCGTGCAGCGTGCAGCGTGCAGCGTGCAGCGTGCAGCGT 1218
QY 1114 GAGGACACCGCGCTGCTGTCAGCGCTGTCAGCGCTGTCAGCGCTGTCAGCGCTGTCAGCG 1173
Db 1219 GAGGACACCGCGCTGCTGTCAGCGCTGTCAGCGCTGTCAGCGCTGTCAGCGCTGTCAGCG 1278
QY 1174 GACAGCAATTAAGCTCAAGCTGCTCTCTCTGATCAATCAATCAATCAATCAATCAATCA 1233
Db 1279 GACAGCAATTAAGCTCAAGCTGCTCTCTCTGATCAATCAATCAATCAATCAATCAATCA 1338
QY 1234 AGCATGGCAGCATCTTACATGCTGCTCTCTCATCAGAAAGCTTTGGAACTATCAAA 1293
Db 1339 AGCATGGCAGCATCTTACATGCTGCTCTCTCATCAGAAAGCTTTGGAACTATCAAA 1398
QY 1294 CTTTCACTACTCTGGAAGGGAACCAAGATTAAGGCACTTCCCTGTTCAAGAACTA 1353
Db 1399 CTTTCACTACTCTGGAAGGGAACCAAGATTAAGGCACTTCCCTGTTCAAGAACTA 1458
QY 1354 TCGAGCAGACTCCCGAAGCCAGTCTGATTAAGAGAGAGCAGACATCCCAAGAGCTG 1413
Db 1459 TCGAGCAGACTCCCGAAGCCAGTCTGATTAAGAGAGAGCAGACATCCCAAGAGCTG 1518
QY 1414 CAGACCGCCAGGCTTCAAGACAGCCAGAGCAAGGATTCATTTGGTCAAGAACAGCAGC 1473
Db 1519 CAGACCGCCAGGCTTCAAGACAGCCAGAGCAAGGATTCATTTGGTCAAGAACAGCAGC 1578
QY 1474 AGTGGCACCGCCAGAGGCTCCTTTATCTCCACATGATGAATGGAGAGCTGGAAGAC 1533
Db 1579 AGTGGCACCGCCAGAGGCTCCTTTATCTCCACATGATGAATGGAGAGCTGGAAGAC 1638
QY 1534 AATTACACACACAGCTTCTTTTGGCTTCCACAGCAGCAGCAGCTCAGCAAGTCT 1593
Db 1639 AATTACACACACAGCTTCTTTTGGCTTCCACAGCAGCAGCAGCTCAGCAAGTCT 1698
QY 1594 GCTGGCTGGGCTTAAAGGCTGCGACTGCGATATCTTGGACCCGCCAGACCTTACCCCT 1653
Db 1699 GCTGGCTGGGCTTAAAGGCTGCGACTGCGATATCTTGGACCCGCCAGACCTTACCCCT 1758
QY 1654 TCCCTGACCGAGCGCTGTTATTTTGGCAGAGGCTCACAATTTACTTCTGCTCAGCC 1713
Db 1759 TCCCTGACCGAGCGCTGTTATTTTGGCAGAGGCTCACAATTTACTTCTGCTCAGCC 1818
QY 1714 ATCTACGAGAGCAGTGCAGTACTCTGCTACAGCTGCAAGCAGCAGCTGCTGCGGA 1773
Db 1819 ATCTACGAGAGCAGTGCAGTACTCTGCTACAGCTGCAAGCAGCAGCTGCTGCGGA 1878
QY 1774 GACCAAGTCTATTTCTGCGCAGGCGCAGAGCCAAAGTGAAGAGCTGCTGCGCGCG 1833
Db 1879 GACCAAGTCTATTTCTGCGCAGGCGCAGAGCCAAAGTGAAGAGCTGCTGCGCGCG 1938
QY 1834 AGCTGGCATGAAGAGAGCCCTTTGAAGACATTTAAAGCAGAAAGCTGCCAAATGGA 1893
Db 1939 AGCTGGCATGAAGAGAGCCCTTTGAAGACATTTAAAGCAGAAAGCTGCCAAATGGA 1998
QY 1894 TTGTGAGAGAGCATCATGTGAGAGAACAGGTCAAGGGAAGAGCTGGGGAAGTGGCAGT 1953

Db	1999	TTTGGAGAGAGCATGTCAGAGAACAGGTCACGGGAAGCTGGGAAAGTGGCAGT	2058
Qy	1954	CAGCTTAGCTTTTCGGGACAGCATGAAATCATTTAGGTCTCC	1995
Db	2059	CAGCTTAGCTTTTCGGGACAGCATGAAATCATTTAGGTCTCC	2100

Search completed: June 22, 2004, 10:19:07
Job time : 3262 secs

This Page Blank (uspto)

- 21 - (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: June 22, 2004, 02:56:17 ; Search time 172 Seconds

(without alignments)
6436.785 Million cell updates/sec

Title: US-10-029-345A-108_COPY_538_2532

Perfect score: 1995
Sequence: 1 atggcccatgagatgatgtg.....tggaaatcatgaggtctcc 1995Scoring table: OLIGO_NTC
Gapop 60.0, Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 0

Total number of hits satisfying chosen parameters: 92308

Minimum DB seq length: 1419

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents NA:*

- 1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
- 2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
- 3: /cgn2_6/prodata/2/ina/6A_COMB.seq:*
- 4: /cgn2_6/prodata/2/ina/6B_COMB.seq:*
- 5: /cgn2_6/prodata/2/ina/PTUS_COMB.seq:*
- 6: /cgn2_6/prodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1893	94.9	1998	4 US-09-816-494-3	Sequence 3, Appl1
2	1893	94.9	3544	4 US-09-816-494-1	Sequence 1, Appl1
3	19	1.0	1830	4 US-09-557-921-1	Sequence 1, Appl1
4	19	1.0	2170	4 US-08-961-527-312	Sequence 312, App
5	19	1.0	2377	4 US-09-920-668-3	Sequence 3, Appl1
6	19	1.0	2387	4 US-08-705-477E-100	Sequence 100, App
7	19	1.0	2653	2 US-08-325-553-1	Sequence 1, Appl1
8	19	1.0	2653	2 US-08-394-152A-1	Sequence 1, Appl1
9	19	1.0	2653	4 US-08-705-477E-1	Sequence 1, Appl1
10	19	1.0	2957	2 US-08-394-152A-48	Sequence 48, Appl1
11	19	1.0	3017	2 US-08-394-152A-39	Sequence 39, Appl1
12	19	1.0	3017	4 US-08-705-477E-39	Sequence 39, Appl1
13	19	1.0	3017	4 US-08-705-477E-128	Sequence 128, App
14	19	1.0	3077	4 US-08-705-477E-90	Sequence 90, Appl1
15	19	0.9	3083	4 US-08-961-527-138	Sequence 138, App
16	18	0.9	3549	4 US-09-008-097-5	Sequence 5, Appl1
17	18	0.9	4046	1 US-07-793-961A-1	Sequence 1, Appl1
18	18	0.9	4046	1 US-08-240-357-1	Sequence 1, Appl1
19	18	0.9	5852	4 US-09-853-768-10	Sequence 10, Appl1
20	18	0.9	7037	4 US-09-853-768-3	Sequence 3, Appl1
21	18	0.9	7900	4 US-08-956-171E-138	Sequence 138, App
22	18	0.9	9751	3 US-09-238-303-7	Sequence 7, Appl1
23	18	0.9	9751	4 US-09-946-239-7	Sequence 7, Appl1
24	18	0.9	19250	4 US-08-961-527-35	Sequence 35, Appl1
25	18	0.9	29485	4 US-09-785-381-6	Sequence 6, Appl1
26	18	0.9	1830121	4 US-09-557-984-1	Sequence 1, Appl1
27	18	0.9	1830121	4 US-09-643-990A-1	Sequence 1, Appl1

C 28	17	0.9	1509	4 US-09-540-236-1860	Sequence 1860, Ap
C 29	17	0.9	1527	4 US-09-489-039A-6404	Sequence 6404, Ap
C 30	17	0.9	1532	4 US-09-459-133-1	Sequence 1, Appl1
C 31	17	0.9	1533	1 US-08-489-733-2	Sequence 2, Appl1
C 32	17	0.9	1533	2 US-08-993-581B-2	Sequence 2, Appl1
C 33	17	0.9	1533	4 US-09-134-078-B	Sequence 8, Appl1
C 34	17	0.9	1596	4 US-09-484-970B-148	Sequence 148, App
C 35	17	0.9	1703	4 US-09-220-132-59	Sequence 59, Appl1
C 36	17	0.9	1856	2 US-08-360-606B-29	Sequence 29, Appl1
C 37	17	0.9	2187	4 US-09-127-219B-2	Sequence 2, Appl1
C 38	17	0.9	2710	4 US-09-620-312D-829	Sequence 829, App
C 39	17	0.9	2789	4 US-09-907-794A-338	Sequence 338, App
C 40	17	0.9	2789	4 US-09-905-125A-338	Sequence 338, App
C 41	17	0.9	2789	4 US-09-902-775A-338	Sequence 338, App
C 42	17	0.9	3175	4 US-09-023-655-1185	Sequence 1185, App
C 43	17	0.9	3250	3 US-08-617-860B-1	Sequence 1, Appl1
C 44	17	0.9	3846	2 US-08-845-161A-5	Sequence 5, Appl1
C 45	17	0.9	3846	3 US-09-270-751-5	Sequence 5, Appl1

ALIGNMENTS

RESULT 1					
US-09-816-494-3					
Sequence 3, Application US/09816494					
Patent No. 6664089					
GENERAL INFORMATION:					
APPLICANT: Meyers, Rachel A.					
TITLE OF INVENTION: 38692 AND 21117, NOVEL DUAL SPECIFICITY					
FILE REFERENCE: 10448-030002					
CURRENT APPLICATION NUMBER: US/09/816,494					
CURRENT FILING DATE: 2001-03-23					
PRIOR APPLICATION NUMBER: US 60/191,858					
PRIOR FILING DATE: 2000-03-24					
NUMBER OF SEQ ID NOS: 10					
SOFTWARE: FastSeq for Windows Version 4.0					
SEQ ID NO 3					
LENGTH: 1998					
TYPE: DNA					
ORGANISM: Homo sapiens					
US-09-816-494-3					
Query Match					
Best Local Similarity 94.9%; Score 1893; DB 4; Length 1998;					
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;					
QY	1				
DB	1				
QY	61	AGTGAACGGAAGAAAGTGTCTTAATTGATAGCCGCAATTTGTGAAATACATATCC			
DB	61	AGTGAACGGAAGAAAGTGTCTTAATTGATAGCCGCAATTTGTGAAATACATATCC			
QY	121	CACATTTGGAAGCCATTAATCACTGCTCAAGCTTATGAGGAAAGTTGCAACG			
DB	121	CACATTTGGAAGCCATTAATCACTGCTCAAGCTTATGAGGAAAGTTGCAACG			
QY	181	GACAAAGTGTATTAACAGAGCTCATCCAGCATTCAGGAAACATAGGTGATGAT			
DB	181	GACAAAGTGTATTAACAGAGCTCATCCAGCATTCAGGAAACATAGGTGATGAT			
QY	241	TGAGTCAGAAAGTTTATGATTAACATCAAGCTCCCAAGATTTCTCTCTTCA			
DB	241	TGAGTCAGAAAGTTTATGATTAACATCAAGCTCCCAAGATTTCTCTCTTCA			
QY	301	GACTGTTTCTCACTGTAATCTTCTGAGTAAGTGAAGAAAGCTTCACTCTGTCAC			
DB	301	GACTGTTTCTCACTGTAATCTTCTGAGTAAGTGAAGAAAGCTTCACTCTGTCAC			
QY	361	CTTGACAGGTGGGTTTGTCTGATTTCTCTGTTGTTCCCTGTGAAAGAAATCC			
DB	361	CTTGACAGGTGGGTTTGTCTGATTTCTCTGTTGTTCCCTGTGAAAGAAATCC			

Db 361 CTTCCAGATGGGTTGGCTGAGTCTCTGTTGTTCCCTGGCCCTCTGTGAAGAAATTC 420
QY 421 ACTGATGCCCTACCTGATTTCTCAGCCTTGTCTACTGTGTGCAACATTGGGCCAAC 480
Db 421 ACTTATGCTCCCTACCTGATTTCTCAGCCTTGTCTACTGTGTGCAACATTGGGCCAAC 480
QY 481 CGAATTTCTCCCAATCTTTATCTTGGCTGCGCAGGAGATGTCCTCAACAAGAGCTGATA 540
Db 481 CGAATTTCTCCCAATCTTTATCTTGGCTGCGCAGGAGATGTCCTCAACAAGAGCTGATG 540
QY 541 CAGAGAAATGGAGTTGGTTATGTGTAAATGCCAGCAATCTGTCCAAAGCCTGACTTT 600
Db 541 CAGAGAAATGGAGTTGGTTATGTGTAAATGCCAGCAATCTGTCCAAAGCCTGACTTT 600
QY 601 ATCCCGAGTCTCATTTCTCCGCTGCTGCTGATGAGAGCTTTTGTGGAATAATTTTG 660
Db 601 ATCCCGAGTCTCATTTCTCCGCTGCTGCTGATGAGAGCTTTTGTGGAATAATTTTG 660
QY 661 CCGTGTGGACAATCAGTAGATTCTATTGAGAAAGCAAAAGCCTCAATGAGATGTT 720
Db 661 CCGTGTGGACAATCAGTAGATTCTATTGAGAAAGCAAAAGCCTCAATGAGATGTT 720
QY 721 CTATGTCATGTTTATGCTGGGATCTCCCGTCCGCCACATCGTATCGCTACATCATG 780
Db 721 CTATGTCATGTTTATGCTGGGATCTCCCGTCCGCCACATCGTATCGCTACATCATG 780
QY 781 AAGAGATGAGCATGCTTTAGATGAGAGCTTACAGATTTGTGAAAGAAAAGACCTACT 840
Db 781 AAGAGATGAGCATGCTTTAGATGAGAGCTTACAGATTTGTGAAAGAAAAGACCTACT 840
QY 841 ATATCTCCAACTTCAATTTTCTGGGCCAACTCTGAGCTATGAGAAAGATTAAAGAC 900
Db 841 ATATCTCCAACTTCAATTTTCTGGGCCAACTCTGAGCTATGAGAAAGATTAAAGAC 900
QY 901 CAGACTGAGCATCAGGGCCAAAGAGCAATCTCACTGTGCACTGTGAGAGCCAAAT 960
Db 901 CAGACTGAGCATCAGGGCCAAAGAGCAATCTCACTGTGCACTGTGAGAGCCAAAT 960
QY 961 GAACTGTCTCTGCTGTCTCAGAGGCTGAGACAGAAAGAGAGCGCCCTCAGTCCACC 1020
Db 961 GAACTGTCTCTGCTGTCTCAGAGGCTGAGACAGAAAGAGAGCGCCCTCAGTCCACC 1020
QY 1021 TGTGCCGATCTGTCTACTCTCAGAGGCAAGCAAAAGGCCCTGTGATCCGCCAGCGTG 1080
Db 1021 TGTGCCGATCTGTCTACTCTCAGAGGCAAGCAAAAGGCCCTGTGATCCGCCAGCGTG 1080
QY 1081 CCCAGGCTGCCAGGCTGAGCGCTGTGAGAGAGAGAGCGCCCTGTGATCCGCCAGCGTG 1140
Db 1081 CCCAGGCTGCCAGGCTGAGCGCTGTGAGAGAGAGAGCGCCCTGTGATCCGCCAGCGTG 1140
QY 1141 AGTGGGCTGACCTGTCTCCAGACAGAGCTGAGAGAGAGATTAAGGCTTCTCTTC 1200
Db 1141 AGTGGGCTGACCTGTCTCCAGACAGAGCTGAGAGAGAGATTAAGGCTTCTCTTC 1200
QY 1201 TCTCTGAGATTAATCAGTTTCAATTTCAATTCAGCCAGATGCGAGATCTTTACATGCTTC 1260
Db 1201 TCTCTGAGATTAATCAGTTTCAATTTCAATTCAGCCAGATGCGAGATCTTTACATGCTTC 1260
QY 1261 TCTCTATCAGAGATGCTTTGAGATTAATTAACAACCTTCACTCTGAGATGAGACCAAC 1320
Db 1261 TCTCTATCAGAGATGCTTTGAGATTAATTAACAACCTTCACTCTGAGATGAGACCAAC 1320
QY 1321 AAGCTATGCGAGTTTCTCCCTGTTTCAGAGAACTATCGAGAGAGACTCCGGAACAGTCTCT 1380
Db 1321 AAGCTATGCGAGTTTCTCCCTGTTTCAGAGAACTATCGAGAGAGACTCCGGAACAGTCTCT 1380
QY 1381 GATTAGAGAGAGAGAGAGAGATCCCAAGAACTGAGACCGCCAGGCTTTGAGACAGCCAG 1440
Db 1381 GATTAGAGAGAGAGAGAGAGATCCCAAGAACTGAGACCGCCAGGCTTTGAGACAGCCAG 1440
QY 1441 AGCAAGCATTTGATTTGGTCTGAGAAAGAGAGAGAGTGCACCGCCAGAGAGTCTCTTTTA 1500
Db 1441 AGCAAGCATTTGATTTGGTCTGAGAAAGAGAGAGAGTGCACCGCCAGAGAGTCTCTTTTA 1500

Db 1441 AGCAAGCATTTGATTTGGTCTGAGAAAGAGAGAGAGTGCACCGCCAGAGAGTCTTTTA 1500
QY 1501 TCTTCACTGATTCGAAGTGGAGGCTGAGAGCAATTAACACACAGCTTCTTTTGGC 1560
Db 1501 TCTTCACTGATTCGAAGTGGAGGCTGAGAGCAATTAACACACAGCTTCTTTTGGC 1560
QY 1561 CTTTCACAGCCACAGACACTCAGAGAGTGTGTGCTGGGCTTTAAGGGCTGGCAC 1620
Db 1561 CTTTCACAGCCACAGACACTCAGAGAGTGTGTGCTGGGCTTTAAGGGCTGGCAC 1620
QY 1621 TCGATATCTTGGCCCGCCAGACCTCTACCCCTTCTGACAGCAGAGCTGGTATTGTC 1680
Db 1621 TCGATATCTTGGCCCGCCAGACCTCTACCCCTTCTGACAGCAGAGCTGGTATTGTC 1680
QY 1681 ACAGAGTCTCAGACTTTCTATCTGTGCTCAGCATCTACGAGAGAGTGCAGTTACTCT 1740
Db 1681 ACAGAGTCTCAGACTTTCTATCTGTGCTCAGCATCTACGAGAGAGTGCAGTTACTCT 1740
QY 1741 GCTTACAGCTGAGAGCTGCGCCACTTGGGAGACCAAGCTATTCTGTGCGGAGCGG 1800
Db 1741 GCTTACAGCTGAGAGCTGCGCCACTTGGGAGACCAAGCTATTCTGTGCGGAGCGG 1800
QY 1801 CAGAAAGCAAGTGAAGAGTGAAGTCTGCGGCGGAGCTGGCATGAGAGAGCCCTTTGAA 1860
Db 1801 CAGAAAGCAAGTGAAGAGTGAAGTCTGCGGCGGAGCTGGCATGAGAGAGCCCTTTGAA 1860
QY 1861 AAGCATTTAAACCGAGAGAGTGCCTCAATGGAATTTGAGAGAGCATCATCTCAGAGAC 1920
Db 1861 AAGCATTTAAACCGAGAGAGTGCCTCAATGGAATTTGAGAGAGCATCATCTCAGAGAC 1920
QY 1921 AGGTACCGGAGAGAGCTGGGAGAGTGGAGTCACTGAGTCTTTGGGAGCATGGA 1980
Db 1921 AGGTACCGGAGAGAGCTGGGAGAGTGGAGTCACTGAGTCTTTGGGAGCATGGA 1980
QY 1981 ATCAATTGAGTCTCC 1995
Db 1981 ATCAATTGAGTCTCC 1995

RESULT 2

US-09-816-494-1
; Sequence 1, Application US/09816494
; Patent No. 6664089
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel A.
; TITLE OF INVENTION: 38692 AND 21117, NOVEL DUAL SPECIFICITY
; FILE REFERENCE: 10448-030002
; CURRENT APPLICATION NUMBER: US/09/816,494
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 60/191,858
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3544
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (589) ... (2583)
US-09-816-494-1

Query Match 94.9%; Score 1893; DB 4; Length 3544;
Best Local Similarity 99.9%; Freq. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGAGCCATGAGATGATGGAATCTCAATTTGTTACTGAGAGGTTGGTGTCTGCTGGA 60
Db 589 ATGAGCCATGAGATGATGGAATCTCAATTTGTTACTGAGAGGTTGGTGTCTGCTGGA 648
QY 61 AGTGAACGGAAGAAAGTGTCTGTTAATTGATAGCGGCAATTTGTGGAATACATATCC 120
Db 61 AGTGAACGGAAGAAAGTGTCTGTTAATTGATAGCGGCAATTTGTGGAATACATATCC 120

649 AGTGNACGAAAAAGTCTGCTAATTGATAGCCGGCCATTGTGGAATACATACATCC 708
121 GCAATTTGGAGCCATTAAATACATCTGCTCAAGCTTATGAGGAGGATTGGACAG 180
709 CACATTTTGGAGCCATTAAATACATCTGCTCAAGCTTATGAGGAGGATTGGACAG 768
181 GACAAAGTTTAAATACAGAGCTCATCCAGCATTCAGGAAACATTAAGTTGACATTGAT 240
769 GACAAAGTTTAAATACAGAGCTCATCCAGCATTCAGGAAACATTAAGTTGACATTGAT 828
241 TGCAGTCAGAGGTTGATGATTAACATCAAGAGCTCCCAAGATGTTGCTCTCTCTTCA 300
829 TGCAGTCAGAGGTTGATGATTAACATCAAGAGCTCCCAAGATGTTGCTCTCTCTTCA 888
301 GACTGTTTCTCACTGATCTTCTGGGTAACCTGGAGAAAGCTTCACTCTGTTCACTG 360
889 GACTGTTTCTCACTGATCTTCTGGGTAACCTGGAGAAAGCTTCACTCTGTTCACTG 948
361 CTTCAGAGTGGGTTTCTGAGTTCTCTCGTGTGTTCCCTGGCCTCTGTGAAGAAATCC 420
949 CTTCAGAGTGGGTTTCTGAGTTCTCTCGTGTGTTCCCTGGCCTCTGTGAAGAAATCC 1008
421 ACTCTAGTCCCTACCTGCAATTTCTAGCCTTGTACTCTGTTCCCAACATTTGGCCAC 480
1009 ACTCTAGTCCCTACCTGCAATTTCTAGCCTTGTACTCTGTTCCCAACATTTGGCCAC 1068
481 CGAATTTCTCCCAATTTTATCTGCTGCTGCGCAGAGATGCTCTCAACAGAGCTGATA 540
1069 CGAATTTCTCCCAATTTTATCTGCTGCTGCGCAGAGATGCTCTCAACAGAGCTGATA 1128
541 CAGCAGAAATGGGATGTTATGTTAAATGACAGCTATACCGTCCAAAGCCTGACTT 600
1129 CAGCAGAAATGGGATGTTATGTTAAATGACAGCTATACCGTCCAAAGCCTGACTT 1188
601 ATCCCCGAGTCTCATTTCTGCGTGTGCTGTGATGACAGCTTTTGTGAGAAATTTTG 660
1189 ATCCCCGAGTCTCATTTCTGCGTGTGCTGTGATGACAGCTTTTGTGAGAAATTTTG 1248
661 CCGTGTGTTGACAAATCAGTAAATTTTCAATTTGAGAAAGCAAAAGCCTCCATGATGTGT 720
1249 CCGTGTGTTGACAAATCAGTAAATTTTCAATTTGAGAAAGCAAAAGCCTCCATGATGTGT 1308
721 CTAGTGCATGTTTACCTGGGATCTCCGCTCCGCCACCATGCGCTACATCATG 780
1309 CTAGTGCATGTTTACCTGGGATCTCCGCTCCGCCACCATGCGCTACATCATG 1368
781 AAGAGATGAGCAATGCTTAAATGAGCTTACAGATTTTGTGAAGAAAGAAAGCTTACT 840
1369 AAGAGATGAGCAATGCTTAAATGAGCTTACAGATTTTGTGAAGAAAGAAAGCTTACT 1428
841 ATATCTCCAAATCTCAATTTTCTGGGCCAATCTCTGAGCTATGAGAAAGATTAAGAAC 900
1429 ATATCTCCAAATCTCAATTTTCTGGGCCAATCTCTGAGCTATGAGAAAGATTAAGAAC 1488
901 CAGACTGGAGCAATCAGAGGCAAAAGCAAACTCAAGCTGCTGACCTGGAAGAGCCAAAT 960
1489 CAGACTGGAGCAATCAGAGGCAAAAGCAAACTCAAGCTGCTGGAAGAGCCAAAT 1548
961 GAACCTGTCTCTGCTCTCTCAGAGGTTGAGACAGAAAGGAGAGAGCCCTCAGTCCACCC 1020
1549 GAACCTGTCTCTGCTCTCTCAGAGGTTGAGACAGAAAGGAGAGAGCCCTCAGTCCACCC 1608
1021 TGTGCGCACTCTGCTAATCTCAGAGGCAAGCAAAAGCCCTGCTGATCCCGCAGCGTG 1080
1609 TGTGCGCACTCTGCTAATCTCAGAGGCAAGCAAAAGCCCTGCTGATCCCGCAGCGTG 1668
1081 CCCAGGTTGCGGCGTGGAGCGGCTGCTTGAAGAGCAAGCCCGTGGTGAACAGGCGTC 1140
1669 CCCAGGTTGCGGCGTGGAGCGGCTGCTTGAAGAGCAAGCCCGTGGTGAACAGGCGTC 1728
1141 AGTGGCTGCACTGTCTCGGAGACAGAGCTGGAAGACAGCAATTAAGCTCAAGCTTCTTC 1200
1729 AGTGGCTGCACTGTCTCGGAGACAGAGCTGGAAGACAGCAATTAAGCTCAAGCTTCTTC 1788

1201 TCTTGATATCAATCAAGTTTCATATTACAGCAGATGAGCAGATCCCTTACATGCTTC 1260
1789 TCTTGATATCAATCAAGTTTCATATTACAGCAGATGAGCAGATCCCTTACATGCTTC 1848
1261 TCTTGATATCAATCAAGTTTCATATTACAGCAGATGAGCAGATCCCTTACATGCTTC 1320
1849 TCTTGATATCAATCAAGTTTCATATTACAGCAGATGAGCAGATCCCTTACATGCTTC 1908
1321 AAGCTATGCAAGTTCTCCCTGTTTCAAGAACTTATGAGAGCAAGCTTCCGAAACCACT 1380
1909 AAGCTATGCAAGTTCTCCCTGTTTCAAGAACTTATGAGAGCAAGCTTCCGAAACCACT 1968
1381 GATTAAGAGAAAGCAGATCCCAAGAGCTGAGACCGGCGGCTTCAAGCAGCAG 1440
1969 GATTAAGAGAAAGCAGATCCCAAGAGCTGAGACCGGCGGCTTCAAGCAGCAG 2028
1441 AGCAAGCATTGCAATTCGCTGAGAAACAGCAGAGTGGCAGCCCGAGAGTCCCTTTTA 1500
2029 AGCAAGCATTGCAATTCGCTGAGAAACAGCAGAGTGGCAGCCCGAGAGTCCCTTTTA 2088
1501 TCTCACTGATCGAAGTGGAGCGGTGAGCAATTAACACACAGCTTCTTTTCCG 1560
2089 TCTCACTGATCGAAGTGGAGCGGTGAGCAATTAACACACAGCTTCTTTTCCG 2148
1561 CTTCACACAGCAGACAGCACTCAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1620
2149 CTTCACACAGCAGACAGCACTCAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2208
1621 TCGGATATCTTGGCCCCCAGACCTTCAACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1680
2209 TCGGATATCTTGGCCCCCAGACCTTCAACCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 2268
1681 ACAGAGTCTCACTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740
2269 ACAGAGTCTCACTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2328
1741 GCTTACAGTGGAGCAGCTGCGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1800
2329 GCTTACAGTGGAGCAGCTGCGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2388
1801 CAGAGCCAAATGAGCAGAGCTGCTGCGCGCAGAGCTGCTGCTGCTGCTGCTGCTGCTG 1860
2389 CAGAGCCAAATGAGCAGAGCTGCTGCGCGCAGAGCTGCTGCTGCTGCTGCTGCTGCTG 2448
1861 AAGCAGTTTAAACGAGAAAGCTGCCAAATGGAATTTGAGAGAGCATCATGTCAAGAAC 1920
2449 AAGCAGTTTAAACGAGAAAGCTGCCAAATGGAATTTGAGAGAGCATCATGTCAAGAAC 2508
1921 AGTCAAGGAAAGAGCTGGGAAAGTGGGAGATCATGCTTTTCCGCGCAGATGGA 1980
2509 AGTCAAGGAAAGAGCTGGGAAAGTGGGAGATCATGCTTTTCCGCGCAGATGGA 2568
1981 ATCATTTAGGCTCC 1995
2569 ATCATTTAGGCTCC 2583

RESULT 3
US-09-557-921-1
Sequence 1, Application us/09557921
Patent No. 6551810
GENERAL INFORMATION:
APPLICANT: Lucite, Ralf M.
APPLICANT: Wei, Bo
TITLE OF INVENTION: DSP-10 DUAL-SPECIFICITY PHOSPHATASE
FILE REFERENCE: 200125.416
CURRENT APPLICATION NUMBER: US/09/557,921
CURRENT FILING DATE: 2000-04-20
NUMBER OF SEQ ID NOS: 20
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 1830

TYPE: DNA
ORGANISM: Homo sapiens
US-09-557-921-1

Query Match
Best Local Similarity 1.0%; Score 19; DB 4; Length 1830;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 745 TCCGCTCCGCCACCATCG 763
DB 1323 TCCGCTCCGCCACCATCG 1341

RESULT 4
US-08-961-527-312
Sequence 312, Application US/08961527
Patent No. 6420135

GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue

CITY: Rockville
STATE: Maryland
COUNTRY: USA

ZIP: 20850
COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527

FILING DATE:
CLASSIFICATION: 424

PRIOR APPLICATION DATA:
APPLICATION NUMBER:

FILING DATE:
ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 312:
SEQUENCE CHARACTERISTICS:

LENGTH: 2170 base pairs
TYPE: nucleic acid

STRANDEDNESS: double
TOPOLOGY: linear

US-08-961-527-312

Query Match
Best Local Similarity 1.0%; Score 19; DB 4; Length 2170;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 456 ACCGTGCGCACCATTTGG 474
DB 580 ACCGTGCGCACCATTTGG 598

RESULT 5
US-09-920-668-3
Sequence 3, Application US/09920668
Patent No. 6482644

GENERAL INFORMATION:
APPLICANT: Lex M. Cowsett

APPLICANT: Brett P. Monia
TITLE OF INVENTION: ANTISENSE MODULATION OF DUAL SPECIFIC PHOSPHATASE 8 EXPRESSION

FILE REFERENCE: RTS-0246

CURRENT APPLICATION NUMBER: US/09/920,668
CURRENT FILING DATE: 2001-08-01
NUMBER OF SEQ ID NOS: 49

SEQ ID NO 3
LENGTH: 2377

TYPE: DNA
ORGANISM: Homo sapiens

FEATURE:
NAME/KEY: CDS
LOCATION: (135) ... (2012)

US-09-920-668-3

Query Match
Best Local Similarity 1.0%; Score 19; DB 4; Length 2377;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 766 ATCGCTACATCATGAGA 784
DB 906 ATCGCTACATCATGAGA 924

RESULT 6
US-08-705-477E-100
Sequence 100, Application US/08705477E

Patent No. 6569432
GENERAL INFORMATION:

APPLICANT: Israeli, Ron S
APPLICANT: Heston, Warren D.W.

APPLICANT: Fair, William R.
APPLICANT: Overfelli, Outhak

APPLICANT: Pinto, John
TITLE OF INVENTION: PROSTATE-SPECIFIC MEMBRANE ANTIGEN AND USES THEREOF

FILE REFERENCE: 1769/41426-G
CURRENT APPLICATION NUMBER: US/08/705,477E

CURRENT FILING DATE: 1996-08-29
NUMBER OF SEQ ID NOS: 128

SOFTWARE: PatentIn version 3.1
SEQ ID NO 100

LENGTH: 2387
TYPE: DNA

ORGANISM: Homo sapiens
US-08-705-477E-100

Query Match
Best Local Similarity 1.0%; Score 19; DB 4; Length 2387;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 AGATGTGCTCTCTCTCT 297
DB 35 AGATGTGCTCTCTCTCT 53

RESULT 7
US-08-325-553-1
Sequence 1, Application US/08325553

Patent No. 5538866
GENERAL INFORMATION:

APPLICANT: Israeli, Ron S.
APPLICANT: Heston, Warren D.W.

APPLICANT: Fair, William R.
TITLE OF INVENTION: THE PROSTATE-SPECIFIC MEMBRANE ANTIGEN

NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza

CITY: New York
STATE: New York

COUNTRY: United States of America
ZIP: 10112

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/325.553
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/973.337A
FILING DATE: 05 NOV 1992
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 1747/41426
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 664-0525
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2653 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: Carcinoma
IMMEDIATE SOURCE:
CLONE: Prostate-Specific Membrane Antigen
FEATURE:
NAME/KEY: CDS
LOCATION: 262..2511
US-08-325-553-1

Query Match 1.0%; Score 19; DB 1; Length 2653;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 AGATGTTGCCTCTCTCT 297
DB 35 AGATGTTGCCTCTCTCT 53

RESULT 8
US-08-394-152A-1
Sequence 1, Application US/08394152A
Patent No. 5935818
GENERAL INFORMATION:
APPLICANT: Israel, Ron S.
APPLICANT: Heaton, Warren D.W.
APPLICANT: Fair, William R.
TITLE OF INVENTION: PROSTATE-SPECIFIC MEMBRANE ANTIGEN AND
TITLE OF INVENTION: USES THEREOF
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM 330 466 DX2
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/394.152A
FILING DATE: 24-FEB-95
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.

REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 41426-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2653 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: Carcinoma
IMMEDIATE SOURCE:
CLONE: Prostate-Specific Membrane Antigen
FEATURE:
NAME/KEY: CDS
LOCATION: 262..2511
US-08-394-152A-1

Query Match 1.0%; Score 19; DB 2; Length 2653;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 AGATGTTGCCTCTCTCT 297
DB 35 AGATGTTGCCTCTCTCT 53

RESULT 9
US-08-705-477E-1
Sequence 1, Application US/08705477E
Patent No. 6569432
GENERAL INFORMATION:
APPLICANT: Israel, Ron S.
APPLICANT: Heaton, Warren D.W.
APPLICANT: Fair, William R.
APPLICANT: Overfell, Quatek
APPLICANT: Pinto, John
TITLE OF INVENTION: PROSTATE-SPECIFIC MEMBRANE ANTIGEN AND USES THEREOF
FILE REFERENCE: 1769/41426-G
CURRENT APPLICATION NUMBER: US/08/705.477E
CURRENT FILING DATE: 1996-08-29
NUMBER OF SEQ ID NOS: 128
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 2653
TYPE: DNA
ORGANISM: Homo sapiens
US-08-705-477E-1

Query Match 1.0%; Score 19; DB 4; Length 2653;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 AGATGTTGCCTCTCTCT 297
DB 35 AGATGTTGCCTCTCTCT 53

RESULT 10
US-08-394-152A-48/C
Sequence 48, Application US/08394152A
Patent No. 5935818
GENERAL INFORMATION:
APPLICANT: Israel, Ron S.
APPLICANT: Heaton, Warren D.W.
APPLICANT: Fair, William R.
TITLE OF INVENTION: PROSTATE-SPECIFIC MEMBRANE ANTIGEN AND

```

; TITLE OF INVENTION: USES THEREOF
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM 330 466 DX2
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/394,152A
; FILING DATE: 24-FEB-95
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 41426-B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2957 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo Sapien
; TISSUE TYPE: Carcinoma
; IMMEDIATE SOURCE:
; CLONE: Prostate Specific Membrane Antigen
; US-08-394-152A-48

Query Match          1.0%; Score 19; DB 2; Length 2957;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      279 AGATGTGCTCTCTCTCT 297
DB      499 AGATGTGCTCTCTCTCT 481

RESULT 11
US-08-394-152A-39
; Sequence 39, Application US/08394152A
; Patent No. 5935818
; GENERAL INFORMATION:
; APPLICANT: Israeli, Ron S.
; APPLICANT: Heston, Warren D.W.
; APPLICANT: Fair, William R.
; TITLE OF INVENTION: PROSTATE-SPECIFIC MEMBRANE ANTIGEN AND
; TITLE OF INVENTION: USES THEREOF
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM 330 466 DX2
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/394,152A
; FILING DATE: 24-FEB-95
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 41426-B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3017 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo Sapien
; TISSUE TYPE: Carcinoma
; IMMEDIATE SOURCE:
; CLONE: Prostate Specific Membrane Antigen
; US-08-394-152A-39

Query Match          1.0%; Score 19; DB 2; Length 3017;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      279 AGATGTGCTCTCTCTCT 297
DB      2519 AGATGTGCTCTCTCTCT 2537

RESULT 12
US-08-705-477E-39
; Sequence 39, Application US/08705477E
; Patent No. 6569432
; GENERAL INFORMATION:
; APPLICANT: Israeli, Ron S.
; APPLICANT: Heston, Warren D.W.
; APPLICANT: Fair, William R.
; APPLICANT: Overfelli, Quathak
; APPLICANT: Pinto, John
; TITLE OF INVENTION: PROSTATE-SPECIFIC MEMBRANE ANTIGEN AND USES THEREOF
; FILE REFERENCE: 1769/41426-G
; CURRENT APPLICATION NUMBER: US/08/705,477E
; CURRENT FILING DATE: 1996-08-29
; NUMBER OF SEQ ID NOS: 128
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 39
; LENGTH: 3017
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-08-705-477E-39

Query Match          1.0%; Score 19; DB 4; Length 3017;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      279 AGATGTGCTCTCTCTCT 297
DB      2519 AGATGTGCTCTCTCTCT 2537

RESULT 13
US-08-705-477E-128
; Sequence 128, Application US/08705477E
; Patent No. 6569432
; GENERAL INFORMATION:
; APPLICANT: Israeli, Ron S
```

APPLICANT: Heston, Warren D.W.
APPLICANT: Fair, William R.
APPLICANT: Overfelli, Ouathek
APPLICANT: Pinto, John
TITLE OF INVENTION: PROSTATE-SPECIFIC MEMBRANE ANTIGEN AND USES THEREOF
FILE REFERENCE: 1769/41426-G
CURRENT APPLICATION NUMBER: US/08/705,477E
CURRENT FILING DATE: 1996-08-29
NUMBER OF SEQ ID NOS: 128
SOFTWARE: PatentIn version 3.1
SEQ ID NO 128
LENGTH: 3017
TYPE: DNA
ORGANISM: Homo sapiens
US-08-705-477E-128

Query Match 1.0%; Score 19; DB 4; Length 3017;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 19; Conservative 0; Indels 0; Gaps 0;

QY 279 AGATGTCCTCTCTCTCT 297
DB 2519 AGATGTCCTCTCTCTCT 2537

RESULT 14
US-08-705-477E-90
Sequence 90, Application US/08/05477E
Patent No. 6569432
GENERAL INFORMATION:
APPLICANT: Israel, Ron S
APPLICANT: Heston, Warren D.W.
APPLICANT: Fair, William R.
APPLICANT: Overfelli, Ouathek
APPLICANT: Pinto, John
TITLE OF INVENTION: PROSTATE-SPECIFIC MEMBRANE ANTIGEN AND USES THEREOF
FILE REFERENCE: 1769/41426-G
CURRENT APPLICATION NUMBER: US/08/705,477E
CURRENT FILING DATE: 1996-08-29
NUMBER OF SEQ ID NOS: 128
SOFTWARE: PatentIn version 3.1
SEQ ID NO 90
LENGTH: 3077
TYPE: DNA
ORGANISM: Homo sapiens
US-08-705-477E-90

Query Match 1.0%; Score 19; DB 4; Length 3077;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 AGATGTCCTCTCTCTCT 297
DB 2579 AGATGTCCTCTCTCTCT 2597

RESULT 15
US-08-961-527-138/C
Sequence 138, Application US/08961527
Patent No. 6420135
GENERAL INFORMATION:
APPLICANT: Charlee Kunach
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage

COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 138:
SEQUENCE CHARACTERISTICS:
LENGTH: 3083 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-138

Query Match 0.9%; Score 18; DB 4; Length 3083;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 951 GAAGCCAATGAACCTGT 968
DB 748 GAAGCCAATGAACCTGT 731

Search completed: June 22, 2004, 09:24:34
Job time: 178 secs

This Page Blank (cspto)

Db 21 SerGlyThrGluValLeuLeuIleAspSerArgProPheValGlnTyrAsnHisSer 40
Qy 121 CACATTTTGGAAAGCCATTAAATACAACTGCTCCAAAGCTTATGAAAGAGTTGGACAG 180
Db 41 HisIleLeuGlnAlaIleAsnIleAsnCysSerLysLeuMetLysArgArgLeuGlnGln 60
Qy 181 GACAAGTGTAAATTAACAGAGCTCATCCAGCATTCAGGAAACCTAAAGTTGACATTGAT 240
Db 61 AspLysValLeuIleThrGluLeuIleGlnHisSerAlaLysHisLysValAspIleAsp 80
Qy 241 TGCAGTCAGAAAGTTGATGATTACGATCAAAAGCTCCCAAGATGTTGCTCTCTCTTCA 300
Db 81 CysSerGlnLysValValValValLysArgLysSerGlnAspValAlaSerLeuSerSer 100
Qy 301 GACGTGTTTTCACACGTACTCTGGGTAAACTGGAGAAAGACCTTCACTGTTACCTG 360
Db 101 AspCysPheLeuThrValLeuLeuGlnLysLeuGlnLysSerPheAsnSerValHisLeu 120
Qy 361 CTTCAGAGTGGGATTGCTGAGTTCTCTGTTGTTTCCCTGGCCCTGTGAAAGAAATCC 420
Db 121 LeuAlaGlyLysPheAlaGlnPheSerArgCysPheProGlyLeuCysGlnGlyLysSer 140
Qy 421 ACTTAAGTCCCTACCTGCAATTTCTCAAGCTTGTACCTGTGTCCAATTTGGGCCAAC 480
Db 141 ThrLeuValProThrCysIleSerGlnProCysLeuProValAlaAsnIleGlyProThr 160
Qy 481 CGAATTCCTCCCAATCTTATCTTGGCTCCAGGAGATGTCCTCAACAGAGCTGATA 540
Db 161 ArgIleLeuProAsnLeuIleLysGlnCysGlnArgAspValLeuAsnLysGlnLeuIle 180
Qy 541 CAGCAGAAATGGGATTTGTTATGTGTAAATGCCAGCTATACCTGTCCAAAGCCTGACTT 600
Db 181 GlnIleAsnGlyIleGlyLysValLeuAsnAlaSerLysThrCysProLysProAspPhe 200
Qy 601 ATCCCGAGTCTACTTCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
Db 201 IleProIleSerHisPheLeuArgValProValAsnAspSerPheCysGlnLysIleLeu 220
Qy 661 CCGTGTGTGACAAATCAGTAGATTTCATTGAGAAAGCAAAAGCCTCCCAATGGATGTT 720
Db 221 ProThrLeuAspLysSerValAspPheIleGlnLysAlaLysAlaSerAsnGlyCysVal 240
Qy 721 CTAGTGCATCTGTTAGCTGGGATCTCCGCTCCGCCAATCGCTATCGCTACATCAGT 780
Db 241 LeuValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIleMet 260
Qy 781 AAGAGATGAGCATGTCTTAGATGAAGCTTACGATTTGTGAAGAAAGAAAGACCTACT 840
Db 261 LysArgMetAspMetSerLeuAspGlnAlaTyrArgPheValLysGlnLysArgProThr 280
Qy 841 ATATCTCAAACTTCAATTTCTGGGCCAACTCTGCACTATGAGAAAGATTAAAGAC 900
Db 281 IleSerProAsnPheAsnPheLeuGlnGlnLeuLeuAspLysGlnLysLysIleLeuAsn 300
Qy 901 CAGACTGGAGCATCAGAGGCCAAAGACAACTAACTGCTGCACTGAGAGAGCCAAAT 960
Db 301 GlnThrGlyAlaSerLysProLysSerLysLeuLysLeuHisIleLeuGlnLysProAsn 320
Qy 961 GAACCTGTCCCTGCTCTCAGAGGGTGGACAGAAAGGAGAGAGCCCTCAGTCCACCC 1020
Db 321 GluProValProAlaValSerGlnGlyGlnLysSerGlnThrProLeuSerProPro 340
Qy 1021 TGTGCGGACTGTGCTACCTCAGAGGAGAGAGCAAAAGGCCGCTGATCCCGCAGCGTG 1080
Db 341 CysAlaAspSerAlaThrSerGlnAlaAlaGlnArgProValHisProAlaSerVal 360
Qy 1081 CCAAGGCTGCCAGGCGTGAGCGCTCGCTGTAGAGAGACAGCCCGCTGTACAGCGCTC 1140
Db 361 ProSerValProSerValGlnProSerLeuLeuGlnAspSerProLeuValGlnAlaLeu 380
Qy 1141 AGTGGCTGCACCTGTCCGACAGAGCTGAGAGACAGCAATTAAGCTTAAGGTTCTTC 1200

Db 381 SerGlyLeuHisLeuSerAlaAspArgLeuGlnAspSerAsnLysLeuLysArgSerPhe 400
Qy 1201 TCTGGATATCAAAATCAGTTTCACTATTCAGCAGAGATGGCAGCATCTTACATGCTTC 1260
Db 401 SerLeuAspIleLysSerValSerLysSerAlaSerMetAlaAlaSerLeuHisGlyPhe 420
Qy 1261 TCTTCATCAGAAAGATCTTTGGAAATCTACAAACCTTCCACTACTGTGATGGAGCAAC 1320
Db 421 SerSerSerGlnAspAlaLeuGlnLysTyrLysProSerThrThrLeuAspGlyThrAsn 440
Qy 1321 AAGCTATGCGAGTTCCTCCCTGTTCAGGAACTATGAGAGCAACATCTCCGAAACCATGCT 1380
Db 441 LysLeuCysGlnPheSerProValGlnGlnLeuSerGlnGlnThrProGlnThrSerPro 460
Qy 1381 GATTAAGAGAAAGCCAGCATCCCAAGAGCTGACAGACCGCAGGCTTACAGACAGCCAG 1440
Db 461 AspLysGlnGlnAlaSerIleProLysLysLeuGlnThrAlaArgProSerAspSerGln 480
Qy 1441 AGCAGCCATTTGCATTCGCTCAGAAACAGACAGCACTGAGCCAGCCGACAGGTCCTTTTA 1500
Db 481 SerLysArgLeuHisSerValArgThrSerSerSerGlyThrAlaGlnArgSerLeuLeu 500
Qy 1501 TCTCCAGTCGATCGAAGTGGGAGCGGTGAGAGCAATTAACACACAGCTTCTTTCCGC 1560
Db 501 SerProLeuHisArgSerGlySerValGlnAspAsnTyrHisThrSerPheLeuPheGly 520
Qy 1561 CTTTTCACAGCAGCAGACACCTCAAGAACTGTGCTGCTGAGCTTAAAGAGCTGAGC 1620
Db 521 LeuSerThrSerGlnGlnHisIleLeuThrLysSerAlaGlyLeuGlyLeuLysGlyTyrHis 540
Qy 1621 TCGGATATCTGGGCCCCCAGACCTTACCCCTTCCCTGACACAGCAGCTGTGATTTGGC 1680
Db 541 SerAspIleLeuAlaProGlnThrSerThrProSerLeuThrSerSerThrLysPheAla 560
Qy 1681 ACAGAGCTCTCACTTCTACTGCTGCTCAGCAGCATTTACGAGAGGACAGTCACTACT 1740
Db 561 ThrGlnSerSerHisPheLysSerAlaSerAlaIleTyrGlyLysSerAlaSerLysSer 580
Qy 1741 GCCTACAGCTGACGACGCTGCCCATTTGGCGAAGCAAGCTTATTTGTGCGGAGCGG 1800
Db 581 AlaTyrSerCysSerGlnLeuProThrCysGlyAspGlnValTyrSerValArgArgArg 600
Qy 1801 CAGAAAGCAATGACAGAGCTGACTCGCGGAGGCTGGCATGAGAGAGAGCCCTTTGAA 1860
Db 601 GlnLysProSerAspArgAlaAspSerArgArgSerLysGlnGlnLysProPheGln 620
Qy 1861 AAGCATTTAAACGACAGAGCTGCCAATGGAATTTGGAGAGAGCATCATGTACAGAAAC 1920
Db 621 LysGlnPheLysArgArgSerCysGlnMetGlnPheGlyGlnSerIleMetSerGlnAsn 640
Qy 1921 AGGTCAAGGAGAGAGCTGGGGAAGTGGGAGTCAGTCTTACGCTTTTGGGACAGATGAA 1980
Db 641 ArgSerArgGlnGlnLeuGlyLysValGlySerGlnSerSerPheSerGlySerMetGln 660
Qy 1981 ATCATTGAGGTCTCC 1995
Db 661 IleIleGlnValSer 665

RESULT 2
ABRS2407
ID ABRS2407 standard; protein; 665 AA.
XX ABRS2407;
DT 19-JUN-2003 (first entry)
XX
XX Protein relating to the invention SEQ ID NO: 148.
XX
XX antiproliferative; hepatotropic; nephrotropic; antiarthritic;
XX antiproliferative; cardiatic; cyostatic; gene therapy; liver disease;
XX proliferative disorder; renal failure; cardiovascular disorder;
XX immunological disorder; arthritis; psoriasis; congenital heart defect;
XX congestive heart failure; cancer; human; mouse; yeast; rat; fruitfly.

XX Homo sapiens.
 OS
 XX
 PN WO200257460-A2.
 XX
 PD 25-JUL-2002.
 XX
 PP 20-DEC-2001; 2001MO-US050459.
 XX
 PR 20-DEC-2001; 2000US-0256668P.
 PR 30-MAR-2001; 2001US-0280186P.
 PR 01-MAY-2001; 2001US-0287735P.
 PR 05-JUN-2001; 2001US-0295448P.
 PR 25-JUN-2001; 2001US-0300465P.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 PI Jackson DG, Feder J, Nelson T, Muntler G, Ramanathan C, Lee L,
 PI Siemers N, Bol D, Schieven G, Finger J, Todderud CG, Bassolino D,
 PI Kyserek S, Mcatee P, Suchard S, Banas D;
 DR MPI; 2002-559721/64.
 XX N-PSDB; ACC60572.
 XX
 PT Novel polynucleotides encoding human phosphatase polypeptides useful in
 PT the prevention or treatment of e.g. proliferative and cardiovascular
 PT disorders.
 XX
 PS Disclosure; Fig 19; 801pp; English.
 XX
 CC The invention relates to a novel isolated nucleic acid comprising a
 CC polynucleotide having a nucleotide sequence selected from 40
 CC polynucleotides fully defined in the specification. The polynucleotide of
 CC the invention has antiproliferative, hepatocytic, nephrotropic,
 CC antitumor, antiproliferative, cardiac, and cytostatic activity. The
 CC polynucleotide may have a use in gene therapy. A polynucleotide or
 CC polypeptide of the invention is useful for preventing, treating or
 CC ameliorating a medical condition, e.g. a proliferative disorder. They are
 CC also useful for treating e.g. liver disease, renal failure, immunological
 CC disorders including arthritis and psoriasis, cardiovascular disorders
 CC such as congenital heart defects and congestive heart failure, and
 CC cancer. A method of the invention is useful for diagnosing a pathological
 CC condition or susceptibility to a condition in a subject. The present
 CC sequence is used in the exemplification of the invention
 CC
 XX
 SQ Sequence 665 AA:
 Alignment Scores:
 Pred. No.: 0 Length: 665
 Score: 665.00 Matches: 665
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 5 Gaps: 0
 US-10-029-345A-108_COPY_538_2532 (1-1995) x ABR52407 (1-665)
 QY 1 ATGGCCCATGATGATGAGTGAACCTCAATGTTACTGAGAGGTTGGCTGCTGCGAA 60
 Db 1 MetAlaHisGluMetIleGlyThrGlnIleValThrIuArgLeuValAlaLeuGlu 20
 QY 61 AGTGAACGGAAGAAAGTGTCTATGATGATGAGCCGCGCATTTGTGGAATCAATACATCC 120
 Db 21 SerGlyThrGluValLeuLeuIleAspSerArgProPheValGluTyraAsnThrSer 40
 QY 121 CACATTTGGAAGCATTAATATCACTGCTCCAGGTTATGGAAGGAGGTTGCAACAG 180
 Db 41 HisIleuGluIuAlaIleAsnIleAsnCysSerIuSerIuValArgGluGln 60
 QY 181 GACAAAGTGTATTAATACAGAGCTCATCAGCATTCAGCGAAACATTAAGTTGACATTGAT 240
 Db 61 AspArgValLeuIleThrGluLeuIleGlnHisSerIuAluYbHisIuValAspIleAsp 80

QY 241 TGCAGTCAGAAAGTTGATTAAGTACATCAAGCTCCAGATGTTGCTCTCTCTCA 300
 Db 81 CysSerGlnIuValValIuTyraAspGlnSerSerGlnAspValAlaSerIuSerSer 100
 QY 301 GACTGTTTCTCAGTGTACTTCTGGGTAACTGGAAGAGGCTTCACTTGTTCACCTG 360
 Db 101 AspCysPheLeuThrValLeuGluGlyIuSerGlnIuValSerPheAsnSerValHisIleu 120
 QY 361 CTTCGACAGTGGGTTGCTGATGTTCTCTGTTGTTCCCTGGCTCTGTGTAAGAAATTC 420
 Db 121 LeuAlaGlyGlyPheAlaGluPheSerArgCysPheProGlyLeuGlyGluGlySer 140
 QY 421 ACTGTAGTCCATCCAGCATTTCTCAGCTGCTTACTGTTGCCAACTGGGCGCAAC 480
 Db 141 ThrLeuValProThrCysIleSerGlnProCysLeuProValAlaAsnIleGlyProThr 160
 QY 481 CGAATTCCTCCAACTTTATCTTGTGGCTCCAGCGATGTCCTCAACAGAGCTGATA 540
 Db 161 ArgIleLeuProAsnLeuIuArgGlnArgAspValLeuAsnIuSerGluLeuIle 180
 QY 541 CAGCAGAAATGGAGTTGTTATGTTAAATGCCAGCTATACCTGTGCCAAAGCTGACTT 600
 Db 181 GlnGlnAsnGlyIleGlyTyraValLeuAsnAlaSerTyraThrCysProIuSerProAspPhe 200
 QY 601 ATCCCGAGTTCATTCCTGCGGTGCTGATGATGATGATGATGATGATGATGATGATG 660
 Db 201 IleProGlnSerHisPheLeuArgValProValAsnAspSerPheCysGluIuSerIleu 220
 QY 661 CCGTGTGAGCAAAATCACTAGATTTCAATTTGAGAAAGCAAAAGCTTCAATGATGTTG 720
 Db 221 ProThrLeuAspIuSerValAlaPheIleGluValAlaValAlaSerAsnGlyCysVal 240
 QY 721 CTATGTCATGTTTATGCTGGATCTCCGCTCCGCCACATGCTATTCGCTATCATCATG 780
 Db 241 LeuValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyriIleMet 260
 QY 781 AAGAGATGAGCAATGCTTTAGTAGAGCTTCAGATTTGTGTAAGAAAGAAAGCACTACT 840
 Db 261 LysArgMetAspSerSerLeuAspGlnAlaTyraArgPheValAlaGluIuSerArgProThr 280
 QY 841 ATATTCCTCAAACTTCAATTTTCTGGGCAACTCTGCTGATGATGAGAAAGATTAAGAAC 900
 Db 281 IleSerProAsnPheAsnIleuGlnIleuLeuAspTyrcIuIuSerIleIuSerAsn 300
 QY 901 CAGACTGAGCATCAGCGCCAAAGAGCAAACTCAAGCTGCTGCACTGGAAGAACCAAT 960
 Db 301 GlnThrGlyAlaSerGlyProIuSerIuSerIuSerIuSerIuSerIuSerIuSerIuSer 320
 QY 961 GAACCTGCTCCGCTGCTCTCAGAGGGTGAAGAAAGGAGCGCCCTCACTCCAGCC 1020
 Db 321 GluProValProAlaValSerGluGlyGlnIuSerGlnIuSerGlnIuSerProIuSerPro 340
 QY 1021 TGTGCGCACTGTGCTACTCAGAGGAGGAGGAGCAAAAGCCCTGATCCCGCAGCGTG 1080
 Db 341 CysAlaAspSerAlaThrSerGluAlaIleGlyGlnArgProValHisProAlaSerVal 360
 QY 1081 CCCAGCGTCCAGCGTGAAGCCGCTGCTTTAGAGACAGCCCGCTGTCAGCGGCTC 1140
 Db 361 ProSerValProSerValGlnProSerLeuLeuGlnAspSerProLeuValGlnAlaLeu 380
 QY 1141 AGTGGGCTCACCTGTCCCGCAGAGGCTGGAAGAGCAAGCAATTAAGCTCAAGCGTCTTC 1200
 Db 381 SerGlyLeuHisIleuSerAlaAspArgLeuGlnAspSerAsnIuSerIuValArgSerPhe 400
 QY 1201 TCTCTGATATCAATCACTTATCATATTCAGCCAGCATGAGCATCTTACATGCTTC 1260
 Db 401 SerLeuAspIleIuSerValSerTyraSerAlaSerMetAlaAlaSerIuHisGlyPhe 420
 QY 1261 TCTCATCAAGAGATCTTTGGAATCTAACAACTTCCATCTCTGATGGAGCAAC 1320
 Db 421 SerSerSerGluIuAspAlaLeuGluTyraTyraProSerThrThrLeuAspGlyTyraAsn 440
 QY 1321 AAGCTATGCCAGTTCCTCCCTGTTCAAGGAAGTATCGAGAGAGCTCCGAAACCAAGTCT 1380

Db 441 |||||LysLeuCySGlnPheSerProValGlnGlnLeuSerGlnGlnThrProGlnThrSerPro 460
QY 1381 GATAGAGGAGGAGGAGGATCCCCAAGAGCTGCGAGCCGCGAGCTTCAGACAGCCAG 1440
Db 461 |||||AspLysGlnGlnValAsnSerIleProLysLeuGlnThrAlaArgProSerAspSerGln 480
QY 1441 AGCAGCGATTGCTTCGGTCAAGACGAGCAGTGGCAGCCGCGAGGCTCCCTTTTA 1500
Db 481 SerLysArgLeuHisSerValArgThrSerSerSerGlyThrAlaGlnArgSerLeuLeu 500
QY 1501 TCTCCACTGTCATCGAAGTGGGAGCGTGGAGGACAATTACACACAGCTTCCTTTTCGC 1560
Db 501 SerProLeuHisArgSerGlySerValGlnAspAsnTyrHisThrSerPheLeuPheGly 520
QY 1561 CTTTCCACGAGCAGCAGCAGCTCAACGAAAGTGTGCTGGCCCTTGAAGGCTGGCAGC 1620
Db 521 LeuSerThrSerGlnGlnHisLeuThrLysSerAlaGlyLeuGlyLeuLysGlyTyrHis 540
QY 1621 TCGGATATCTTGAGCCCCCAGACCTTACCCCTTCCCTGACACGAGCAGCTGGTATTTGGC 1680
Db 541 SerAspIleLeuAlaProGlnThrSerThrProSerLeuThrSerSerTyrPheAla 560
QY 1681 ACAGAGTCTTCAACTTCTACTGCTGCTCAGCCATCTTACGAGGAGCAGTGCAGTTACTCT 1740
Db 561 ThrGlnSerSerHisAspPheTyrSerAlaSerAlaIleTyrGlySerAlaSerTyrSer 580
QY 1741 GCTTACAGCTGCGAGCAGCTGCGAGCTGCGAGCAGCAGCAAGCTTATCTGTGCGCAGGCGG 1800
Db 581 AlaTyrSerCysSerGlnLeuProThrCysGlnAspGlnValTyrSerValArgArgArg 600
QY 1801 CAGAGGCCAAGTGAACAAGCTGACTGCGGCGGAGAGCTGGCATGAAGAGAGCCCTTTGAA 1860
Db 601 GlnLysProSerAspArgAlaAspSerArgArgSerThrHisGlnGlnLeuSerProPheGln 620
QY 1861 AAGCAGTTTAAAGCAGAAAGCTGCCAATGGAATTTGAGAGAGCATCATGTCAAGAGAAC 1920
Db 621 LysGlnPheLysArgArgSerCysGlnMetGlnPheGlyGlnSerIleMetSerGlnAsn 640
QY 1921 AGGTCAAGGAGAGAGCTGGGAGAAAGTGGGAGTCAAGTCTAGCTTTTGGGCGAGCATGGAA 1980
Db 641 ArgSerArgGlnGlnLeuGlyLysValGlySerGlnSerSerPheSerGlySerMetGln 660
QY 1981 ATCATTGAGGTCTCC 1995
Db 661 IleIleGlnValSer 665

Search completed: June 21, 2004, 14:24:44
Job time : 100.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 21, 2004, 14:20:20 ; Search time 34.5 Seconds
(without alignments)
11124.759 Million cell updates/sec

Title: US-10-029-345a-108_COPY_538_2532

Perfect score: 665
Sequence: 1 atggcccatcagatgatctgg.....tgggaatcattgagctctcc 1995

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Word size: 473

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 75 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool_p/US10029345/runat_21062004_140535_16051/app_query.fasta_1.2183
-DB=PIR_78 -QFMT=faetan -SUFFIX=0112.rpr -MINMATCH=0.1 -LOOPL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=75
-DOCALIGN=200 -THR_SCORE=quality -THR_MIN=473 -ALIGN=50 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10029345_QCGN_1_1_52@runat_21062004_140535_16051 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : PIR_78:*

1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description

No matches found

Search completed: June 21, 2004, 14:30:35
Job time : 34.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 21, 2004, 14:19:10 ; Search time 22 Seconds

(without alignments)
9443.633 Million cell updates/sec

Title: US-10-029-345a-108_COPY_538_2532

Perfect score: 665
Sequence: 1 atggcccatgagatgattg9.....tggaaatcattgaggtctcc 1995

Scoring table:

OLIGO	Xgapop	Xgapext
60.0	60.0	60.0
Ygapop	60.0	Ygapext
60.0	60.0	60.0
Fgapop	6.0	Fgapext
7.0	6.0	7.0
Delop	6.0	Delext
7.0		

Searched: 141681 seqs, 52070155 residues

Word size: 473

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 75 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp
-O=/cgn2.1/USPTO_spool_p/US10029345/runat_21062004_140534_16025/app_query.fasta_1.2183
-DB=SwissProt 42 -QFMT=fastan -SUFFIX=0112.rsp -MINMATCH=0.1 -LOOPCT=0
-LOOPEXT=0 -UNITS=bites -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=75 -DOCALLIGN=200 -THR SCORE=quality -THR MIN=473 -ALIGN=50 -MODE=LOCAL
-OUTFMT=ptco -NORM=ext -HEADSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USRR=US10029345_@CGN_1_1_22_@runat_21062004_140534_16025 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	ID	Description
No.						

No matches found						

Search completed: June 21, 2004, 14:25:35
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

No matches found

Search completed: June 21, 2004, 14:29:13
Job time : 102.5 secs

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 21, 2004, 14:19:45 ; Search time 102.5 Seconds

(without alignments)

12282.123 Million cell updates/sec

Title: US-10-029-345A-108_COPY_538_2532

Perfect score: 665

Sequence: 1 atggcccatgagatgattg.....tggaatcatgaggtctcc 1995

Scoring table:

OLIGO	
Xgapop 60.0 , Xgapext 60.0	
Ygapop 60.0 , Ygapext 60.0	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 1017041 seqs, 315518202 residues

Word size: 473

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 75 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xlp
-O=/cgn2.1/USPTO_spool_p/US10029345/runat.21062004.140534.16037/app.query.fasta_1.2183
-DB=SPTRMBL.25 -OFMT=fastan -SUPR=ol12.rpt -MISMATCH=0.1 -LOOPCIT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=75 -DOCALLIGN=200 -THR_SCORE=quality -THR_MIN=473 -ALIGN=50 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10029345_@CGN_1.1.112_@runat.21062004.140534.16037 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FCGAPOP=6
-FCGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

SPTRMBL_25: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_phage: *
10: sp_plant: *
11: sp_rodent: *
12: sp_virus: *
13: sp Vertebrate: *
14: sp Unclassified: *
15: sp_rv1rus: *
16: sp_bacteriap: *
17: sp_archaeap: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	DB	ID	Description
--------	-------	-------	-------	--------	----	----	-------------

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 21, 2004, 14:24:51 ; Search time 93.5 Seconds

(without alignments)
12047.395 Million cell updates/sec

Title: US-10-029-345a-108_COPY_538_2532

Perfect score: 665
Sequence: 1 atggccatcagatgattgg.....ttggaatcattgagctctc 1995

Scoring table:

OLIGO	
Xgapop 60.0 , Xgapext 60.0	
Ygapop 60.0 , Ygapext 60.0	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 1163542 seqs, 282313646 residues

Word size: 473

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 75 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=x1p
-Q=/cgn2_1/USPRO.spool_p/US10029345/runat_21062004.140536.16108/app_query.fasta.1.2183
-DB=Published Applications AA -QPMT=fastan -SUFFIX=0112.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo
-TRANS=human40.cdi -LIST=75 -DOCALIGN=200 -THR SCORE=quality -THR MIN=473
-ALIGN=50 -MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10029345_@CGN_1_1_13_@runat_21062004.140536.16108
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THRADS=1 -XGAPOP=60 -XGAPEXT=60
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Published Applications_AA:*

1:	/cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pcp:*
2:	/cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pcp:*
3:	/cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pcp:*
4:	/cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pcp:*
5:	/cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pcp:*
6:	/cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pcp:*
7:	/cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pcp:*
8:	/cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pcp:*
9:	/cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pcp:*
10:	/cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pcp:*
11:	/cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pcp:*
12:	/cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pcp:*
13:	/cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pcp:*
14:	/cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pcp:*
15:	/cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pcp:*
16:	/cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pcp:*
17:	/cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pcp:*
18:	/cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pcp:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match Length	ID	Description
---------------	-----------------------	----	-------------

No matches found

Search completed: June 21, 2004, 14:34:57
Job time : 93.5 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_nzp model

Run on: June 21, 2004, 14:21:20 ; Search time 24.5 Seconds

(without alignments)
8407.658 Million cell updates/sec

Title: US-10-029-345A-108_COPY_538_2532

Perfect score: 665
Sequence: 1 atggcccatgagatgattgg.....tggaaatcatgtgaggtctcc 1995

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 389414 seqs, 51625971 residues

Word size: 473

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 75 summaries

Command line parameters:

-MODEL=frame+nzp.model -DEV=xlp
-O=/cgn2_1/USPTO_spool_p/US10029345/runat_21062004_140535_16081/app_query.fasta_1.2183
-DB=Issued_Patents_AA -OPMT=fastran -SUFFIX=0112.ra1 -MINMATCH=0.1 -IOPCIL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=oligo -TRANS=humand0.cdi
-LIST=75 -POCALIGN=200 -THR_SCORE=quality -THR_MIN=473 -ALIGN=50 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10029345_0CGN_1_1_27@runat_21062004_140535_16081 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEJOURNRY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA:*

1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result	Query	ID	Description
No.	Score	Match length	DB ID

No matches found			

Search completed: June 21, 2004, 14:31:36
Job time : 24.5 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2004, 07:13:40 ; Search time 12404 Seconds

(without alignments)
6971.086 Million cell updates/sec

Title: US-10-029-345A-108_COPY_538_2532

Perfect score: 1995
Sequence: 1 atggcccatgagatgcttgg.....tggaaatcattgaggtctcc 1995

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size : 1419

Total number of hits satisfying chosen parameters: 19

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 75 summaries

Database :

GenEmbl: 1: gb_ba: 2: gb_hcg: 3: gb_in: 4: gb_om: 5: gb_ov: 6: gb_pat: 7: gb_ph: 8: gb_pl: 9: gb_pr: 10: gb_ro: 11: gb_sbs: 12: gb_sy: 13: gb_un: 14: gb_vl: 15: gb_da: 16: em_hum: 17: em_hum: 18: em_mu: 19: em_mu: 20: em_om: 21: em_or: 22: em_ov: 23: em_pac: 24: em_ph: 25: em_pl: 26: em_ro: 27: em_sbs: 28: em_un: 29: em_vl: 30: em_hcg_hum: 31: em_hcg_hum: 32: em_hcg_other: 33: em_hcg_mus: 34: em_hcg_pln: 35: em_hcg_rtd: 36: em_hcg_mam: 37: em_hcg_vrt: 38: em_by: 39: em_hcgo_hum: 40: em_hcgo_mus: 41: em_hcgo_other:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1995	100.0	5450	6	AX482439	AX482439 Sequence
2	1995	100.0	5450	6	AX482439	AX482439 Sequence
3	1893	94.9	3059	6	AX260342	AX260342 Sequence
4	1893	94.9	3059	6	AX278461	AX278461 Sequence
5	1893	94.9	3496	6	AX441210	AX441210 Sequence
6	1893	94.9	3521	6	AB052156	AB052156 Homo sapi
7	1893	94.9	3544	6	AX260340	AX260340 Sequence
8	1893	94.9	3566	9	AF506796	AF506796 Homo sapi
9	1893	94.9	3766	6	AX374994	AX374994 Sequence
10	1893	94.9	4790	6	BD171157	BD171157 Novel gen
11	1893	94.9	4790	6	BD183422	BD183422 Novel gen
12	1893	94.9	4790	6	AB051487	AB051487 Homo sapi
13	1842	92.3	2732	6	AX180875	AX180875 Sequence
14	1842	92.3	3104	6	AX405700	AX405700 Sequence
15	1791	89.8	2102	6	AX713989	AX713989 Sequence
16	1791	89.8	2102	9	AK055973	AK055973 Homo sapi
17	1791	89.8	5111	6	AX482372	AX482372 Sequence
18	1696	85.0	2807	9	BC031643	BC031643 Homo sapi
19	1593	79.8	2118	6	AX099933	AX099933 Sequence

ALIGNMENTS

RESULT 1
AX482439
LOCUS AX482439
DEFINITION Sequence 108 from Patent WO02057460.
ACCESSION AX482439
VERSION AX482439.1 GI:22316984
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Todderud, C.G., Bol, D., Finger, J., Lee, L., Nelson, T., Schieven, G.,
Suchard, S., Banae, D., Bassolino, D., Feder, J., Kytsek, S.,
Mcatee, P., Mintier, G., Siemert, N., Jackson, D.G. and Ramanaathan, C.
TITLE Polynucleotides encoding human phosphatases
JOURNAL Patent: WO 02057460-A 108 25-JUL-2002;
SQUIBB BRISTOL MYERS CO (US)
FEATURES
source
1. 5450
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
538..2535
/note="unassigned protein product"
/codon_start=1
/protein_id="CAD44451.1"
/db_xref="GI:22316985"
/translation="MAHEMIGIOIVTERVALLSEGTREKVLIDSRPFVENVNTHLE
ALININCSKLMKRLQODKVLITELIHSKAKHVDICQKVVYDDSSQDVASLSDC
FLTVLLGKLEKSFNSVHLAAGFAEFSRCPGLCEKSTLVPCISQPLPVANISPT
RILPMLYLGCDRDLVANKELIQNGIGYVNASYTCRPFIPESHLRVPVDSFCEK
ILPWLDSVDFIEKAKASNGCVLHCLAGISSATIAIYIMKMDMSIDEAYRKYKE
KRPTTSPNFPILGOLLDYDEKXIKNOTGASGPKSLKHLKRENEPVPVBSGQKSE
TSLSPCADSATSSEAGRCRPHAPSVSPVQPSLEDSPLVQALSGHLASDRLED
SNLKRKSFSLDIKSVISASMASLSESSBDALRYTPSTLDTNKLCQSPYQOE
LBEQTPETSPDKKEASIPKLTQPARPSDQSKRLHSVRTSSGTAORSLSPHRSHP
VEDNQTPETSPDKKEASIPKLTQPARPSDQSKRLHSVRTSSGTAORSLSPHRSHP
YSASAIYGGASAYSACSGOLPTCCDQVSVRRORQPSDRADSRMRHBSPEKOPK
RSCQMEFGESINSESRREBLKVGSGSFSQSMELIEVS"

ORIGIN

Query Match 100.0%; Score 1995; DB 6; Length 5450;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1995; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGCCCTGAGATGATGAACTCAAAATTTGTACTGAGAGTGTGCTCTGCTGAA 60
Db 538 ATGGCCCTGAGATGATGAACTCAAAATTTGTACTGAGAGTGTGCTCTGCTGAA 597
QY 61 AGTGAACGGAAGAGTCTGCTAATGATAGCCGGCCATTTTGGAAATACAAATACC 120
Db 598 AGTGAACGGAAGAGTCTGCTAATGATAGCCGGCCATTTTGGAAATACAAATACC 657
QY 121 CACATTTTGAAGCCATTATATATCAACTGCTCCAAGCTTATGAGCGAGAGTTGCAAC 180
Db 658 CACATTTTGAAGCCATTATATATCAACTGCTCCAAGCTTATGAGCGAGAGTTGCAAC 717
QY 181 GACAAAGTGTATTAACAGAGCTCATCCAGATTACGCAAAACATAAGTTGACATTGAT 240
Db 718 GACAAAGTGTATTAACAGAGCTCATCCAGATTACGCAAAACATAAGTTGACATTGAT 777
QY 241 TGCAGTCGAAAGTGTAGTTTACGATCAAAAGTCCCAAGATGTGCTCTCTCTCA 300
Db 778 TGCAGTCGAAAGTGTAGTTTACGATCAAAAGTCCCAAGATGTGCTCTCTCTCA 837
QY 301 GACTGTTTCTCACTGACTTCTGGGTAACTGGAAGAGCTTCAACTCTGTTCACTG 360
Db 838 GACTGTTTCTCACTGACTTCTGGGTAACTGGAAGAGCTTCAACTCTGTTCACTG 897
QY 361 CTTCGAGTGGGTTGCTGAGTCTCTGTTGTTTCCCTGGCTCTGTGAGGAAATCC 420
Db 898 CTTCGAGTGGGTTGCTGAGTCTCTGTTGTTTCCCTGGCTCTGTGAGGAAATCC 957
QY 421 ACTCTAGTCCCTCACTGCAATTTCTCAGCCTTGCTTCACTGTTGCAAACTTGGCCCAACC 480
Db 958 ACTCTAGTCCCTCACTGCAATTTCTCAGCCTTGCTTCACTGTTGCAAACTTGGCCCAACC 1017
QY 481 CGAATTTCTCCCAATCTTTATCTTGGCTGCGCAGAGATGTCTCTCAACAGAGCTGATA 540
Db 1018 CGAATTTCTCCCAATCTTTATCTTGGCTGCGCAGAGATGTCTCTCAACAGAGCTGATA 1077
QY 541 CAGCAGAAATGGAGTATGTTATGTTAAATGCCATATACCTGTCCAAAGCTTGACTTT 600
Db 1078 CAGCAGAAATGGAGTATGTTATGTTAAATGCCATATACCTGTCCAAAGCTTGACTTT 1137
QY 601 ATCCCGAGTCTCATTTCTGCGTGTGCTGCTGATGATGACGCTTTTGTGAGAAATTTTG 660
Db 1138 ATCCCGAGTCTCATTTCTGCGTGTGCTGCTGATGATGACGCTTTTGTGAGAAATTTTG 1197
QY 661 CCGTGTGTGGAACAATCAGTAGATTTCATTGAGAAAGCAAAAGCCTCCAAATGATGTGT 720
Db 1198 CCGTGTGTGGAACAATCAGTAGATTTCATTGAGAAAGCAAAAGCCTCCAAATGATGTGT 1257
QY 721 CTAGTCACTGTTTAGCTGGATCTCCGGCTCCGCAACATGGCTATGCGCTTACATCTG 780
Db 1258 CTAGTCACTGTTTAGCTGGATCTCCGGCTCCGCAACATGGCTATGCGCTTACATCTG 1317
QY 781 AAGAGATGACATGTCTTTAGATGAGCTTACAGATTTGTGAAGAAAAAGACCTTACT 840
Db 1318 AAGAGATGACATGTCTTTAGATGAGCTTACAGATTTGTGAAGAAAAAGACCTTACT 1377
QY 841 AATATCTCCCAACTTCAATTTTCTGGGCCAACTCTCTGACATATGAGAAAGATTAGAAC 900
Db 1378 AATATCTCCCAACTTCAATTTTCTGGGCCAACTCTCTGACATATGAGAAAGATTAGAAC 1437
QY 901 CAGACTGAGCATCAGGGCCAAAGAGCAAACTCAAGCTGTGCACTGGAAGAAAGTTAAGAA 960
Db 1438 CAGACTGAGCATCAGGGCCAAAGAGCAAACTCAAGCTGTGCACTGGAAGAAAGTTAAGAA 1497
QY 961 GAACTGTCTCTGCTGTCTCAGAGGTTGACAGAAAAGAGAGAGCCCTCACTCAAGTCC 1020
Db 1498 GAACTGTCTCTGCTGTCTCAGAGGTTGACAGAAAAGAGAGAGCCCTCACTCAAGTCC 1557

QY 1021 TGTGCCGACTGTGCTACTCTCAAGAGCAGAGCAAAAGCCCGTGCATCCGCCAGCGTG 1080
Db 1558 TGTGCCGACTGTGCTACTCTCAAGAGCAGAGCAAAAGCCCGTGCATCCGCCAGCGTG 1617
QY 1081 CCAGAGGTCCAGAGGTGACGCGTGTGTTAGAGGACAGCCCGCTGTACAGCGCTTC 1140
Db 1618 CCAGAGGTCCAGAGGTGACGCGTGTGTTAGAGGACAGCCCGCTGTACAGCGCTTC 1677
QY 1141 AGTGGCTGCACTGTTCGACAGAGCTGGAAGACAGCAATTAAGCTCAAGGTTCTTC 1200
Db 1678 AGTGGCTGCACTGTTCGACAGAGCTGGAAGACAGCAATTAAGCTCAAGGTTCTTC 1737
QY 1201 TCTCTGATATTAATCAAGTTTCAATTTCAAGCTGAGAGCATCTTTAATGCTTC 1260
Db 1738 TCTCTGATATTAATCAAGTTTCAATTTCAAGCTGAGAGCATCTTTAATGCTTC 1797
QY 1261 TCTCATCAAGAAATGCTTTTGAATACAAACCTTGCATCACTGATGGATGGAGCAAC 1320
Db 1798 TCTCATCAAGAAATGCTTTTGAATACAAACCTTGCATCACTGATGGATGGAGCAAC 1857
QY 1321 AAGCTATCCAGTTCTCCCTGTTCAGAGACTATCGAGAGCACTCCGAAACCAAGTCT 1380
Db 1858 AAGCTATCCAGTTCTCCCTGTTCAGAGACTATCGAGAGCACTCCGAAACCAAGTCT 1917
QY 1381 GATTAGAGGAAGCCAGATCCCAAGAACTGCAAGCCGCAAGCTTTCAGACAGCCAG 1440
Db 1918 GATTAGAGGAAGCCAGATCCCAAGAACTGCAAGCCGCAAGCTTTCAGACAGCCAG 1977
QY 1441 AGCAAGCATATTGCAATCGGTGAGAAACAGACAGAGTGGCAACCGCCAGAGGTCCCTTTA 1500
Db 1978 AGCAAGCATATTGCAATCGGTGAGAAACAGACAGAGTGGCAACCGCCAGAGGTCCCTTTA 2037
QY 1501 TCTGCACTGCAATCGAATGGGAGCGTGAAGCAATTAACAACACAGACTTCTTTTGGC 1560
Db 2038 TCTGCACTGCAATCGAATGGGAGCGTGAAGCAATTAACAACACAGACTTCTTTTGGC 2097
QY 1561 CTTCACACAGCCAGCAGACCTCAGAGATCTGCGCTGAGCCCTTAAGGCTGTGCAC 1620
Db 2098 CTTCACACAGCCAGCAGACCTCAGAGATCTGCGCTGAGCCCTTAAGGCTGTGCAC 2157
QY 1621 TCGGATATCTTGGCCCCCAGACCTTACCCCTTCCCTGACCAAGCTGTATTTTGGC 1680
Db 2158 TCGGATATCTTGGCCCCCAGACCTTACCCCTTCCCTGACCAAGCTGTATTTTGGC 2217
QY 1681 ACAGAGTCTTCACTCTTCTACTCTGCTCAGCCATCTACGAGGAGAGTGCAGTTACTCT 1740
Db 2218 ACAGAGTCTTCACTCTTCTACTCTGCTCAGCCATCTACGAGGAGAGTGCAGTTACTCT 2277
QY 1741 GCCTACAGCTGACAGCCAGCTGCCCACTTGCGAGAGCCAAAGTCTATTCTGCGCAGGCGG 1800
Db 2278 GCCTACAGCTGACAGCCAGCTGCCCACTTGCGAGAGCCAAAGTCTATTCTGCGCAGGCGG 2337
QY 1801 CAGAAAGCAAGTACAGAGCTGACTCGCGGCGAGCTGCAATGAAGAGCCCTTTGAA 1860
Db 2338 CAGAAAGCAAGTACAGAGCTGACTCGCGGCGAGCTGCAATGAAGAGCCCTTTGAA 2397
QY 1861 AAGCAGTTTAAAGCAGAAAGCTGCCAAATGGAATTTGAGAGAGCATCATGTCAAGAAC 1920
Db 2398 AAGCAGTTTAAAGCAGAAAGCTGCCAAATGGAATTTGAGAGAGCATCATGTCAAGAAC 2457
QY 1921 AGGTCAACGGAAGAGCTGGGGGAAAGTGGCAAGTCACTAGCTTTTGGGAGAGCAATGAA 1980
Db 2458 AGGTCAACGGAAGAGCTGGGGGAAAGTGGCAAGTCACTAGCTTTTGGGAGAGCAATGAA 2517
QY 1981 ATCATTGAGGTCTCC 1995
Db 2518 ATCATTGAGGTCTCC 2532

RESULT 2
AX482478
LOCUS AX482478 5450 bp DNA linear PAT 16-AUG-2002

DEFINITION Sequence 147 from Patent WO02057460.
ACCESSION AX482478
VERSION AX482478.1 GI:22316999
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Toddernud,C.G., Bol,D., Finger,J., Lee,L., Nelson,T., Schieven,G., Suchard,S., Banas,D., Baesolino,D., Feder,J., Krystek,S., Moatee,P., Minter,G., Siemers,N., Jackson,D.G. and Ramanathan,C.
TITLE Polynucleotides encoding human phosphatases
JOURNAL Patent: WO 02057460-A 147 25-JUL-2002;
SQUIBB BRISTOL MYERS CO (US)
FEATURES
source
1. 5450
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
538..2535
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD44453.1"
/db_xref="GI:22317000"
/db_xref="REFSEQ:CAD44453"
/translation="MAHEMIGTIVTRVALLBSGTEKYLIDSRPVEYNTSHILE
AININCSKLMKRLQODKVLITELIQHSAKVDIDCSQVYVDDSDVDASLSDC
FLTVLGLKLEKSPNSVHLAAGFAEFSRCPGLCEGSTLPTPCICPCLPVANIGPT
RLPNLYLGGORDVNLKELIQNGIGVLAASYCPEDPIPESHFLRYVNDSPCEK
ILPMLKSVDFIEKAVANGCVLHICAGISRSATLAIAYIMRMMSIDEAVPYKE
KRPTISPNRPLQQLIDYEKIKNOTGASPRKSKLHLHEKNEVPVANSBEGQSE
TLPSPCADSAISEAAGQRPVHPASVPSPVPSLLEDSPLVQALSGLHLSADRLD
SNLKXGFSFLDIKSVSYASMAASLHGFSSSEDLERYKSTLLDQTNKLCQSPVQ
LSQETPSTSPDKKEASIPKLTQAPRPSDSQSKLHSVRTSSGTAQRSLSPHRSQS
VEDNYHSFLFGSTSOQLTKSAGLKKMHSDILAPQTSSTLSMTYFATESHP
YSASAIYGSASYSAYSCSOLPTGPOVYSVRRORRDPDADRMRHMBESPREKQFK
RSCQMFEGSINSRRELBKVGSGSFSMSIIEVS"
ORIGIN
Query Match 100.0%; Score 1995; DB 6; Length 5450;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1995; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGCCATGATGATGATGAACTCAATTGTTACTGAGAGTTGGTGGCTCTGCTGAA 60
DB 538 ATGGCCATGATGATGATGAACTCAATTGTTACTGAGAGTTGGTGGCTCTGCTGAA 597
QY 61 AGTGAACGGAAAGGCTGCTAATTGATAGCCGGCATTGTGGAATACATATACATCC 120
DB 598 AGTGAACGGAAAGGCTGCTAATTGATAGCCGGCATTGTGGAATACATATACATCC 657
QY 121 CACATTTTGAAGCCATTATATCACTGCTCCAGCTTATGAAGCGAAGTTGCAACAG 180
DB 658 CACATTTTGAAGCCATTATATCACTGCTCCAGCTTATGAAGCGAAGTTGCAACAG 717
QY 181 GACAAAGTGAATTAATCAGAGCTCATCAGACATTACGGAACATTAAGTTGACATTGAT 240
DB 718 GACAAAGTGAATTAATCAGAGCTCATCAGACATTACGGAACATTAAGTTGACATTGAT 777
QY 241 TGACATGCAAGAGTTGATTAAGTACATCAAGTCCCAAGATGTGGCTCTCTCTTCA 300
DB 778 TGACATGCAAGAGTTGATTAAGTACATCAAGTCCCAAGATGTGGCTCTCTCTTCA 837
QY 301 GACTGTTTCTCACTGACTTCTGGGTAATCGAGAGAAGCTTCACTCTGTTCACTG 360
DB 838 GACTGTTTCTCACTGACTTCTGGGTAATCGAGAGAAGCTTCACTCTGTTCACTG 897
QY 361 CTTCAGAGTGGGTTTGTGAGGTTCTCTGTTGTTTCTCGGCTCTGTGAAGAAAATCC 420
DB 898 CTTCAGAGTGGGTTTGTGAGGTTCTCTGTTGTTTCTCGGCTCTGTGAAGAAAATCC 957
QY 421 ACTCTACTCCCTACCTGCAATTTCTCAGCCTTGTCTTACTCTGTGCAACATTTGGGCAAC 480

DB 958 ACTCTACTCCCTACCTGCAATTTCTCAGCCTTGTCTTACTCTGTGCAACATTTGGGCAAC 1017
QY 481 CGAATTTCTCCCAATTTTATCTCTGGCTGCCAGCGAGATGCTCTCAACAGAGCTGATA 540
DB 1018 CGAATTTCTCCCAATTTTATCTCTGGCTGCCAGCGAGATGCTCTCAACAGAGCTGATA 1077
QY 541 CAGCAATAGGGGTTGGTATGTTAATGACAGCTATACCTGTCCAAAGCCTGACTT 600
DB 1078 CAGCAATAGGGGTTGGTATGTTAATGACAGCTATACCTGTCCAAAGCCTGACTT 1137
QY 601 ATCCCGAGTCTCATTTCTCGGTGTCCTGCTGATGATGACAGCTTTGTGAGAAAATTTG 660
DB 1138 ATCCCGAGTCTCATTTCTCGGTGTCCTGCTGATGATGACAGCTTTGTGAGAAAATTTG 1197
QY 661 CCGTGTGACAAATCAGTATGATTTCAATTGAGAAAGCAAAAGCCTCCAAATGATGTT 720
DB 1198 CCGTGTGACAAATCAGTATGATTTCAATTGAGAAAGCAAAAGCCTCCAAATGATGTT 1257
QY 721 CTAGTGCATGTTTACCTGGGATCTCCGCTCCGACCATGCTATCGCTTACATCATG 780
DB 1258 CTAGTGCATGTTTACCTGGGATCTCCGCTCCGACCATGCTATCGCTTACATCATG 1317
QY 781 AAGAGATGACATGCTTTAGATGAAGCTTACAGATTTGTGAAAAGAAAAGACCTACT 840
DB 1318 AAGAGATGACATGCTTTAGATGAAGCTTACAGATTTGTGAAAAGAAAAGACCTACT 1377
QY 841 ATATCTCCAAATTTCTGAGGCAATCTCTGAGTATGAGAAAGATTAAGAAC 900
DB 1378 ATATCTCCAAATTTCTGAGGCAATCTCTGAGTATGAGAAAGATTAAGAAC 1437
QY 901 CAGCTGAGCATCAGAGGCAAAAGCAAACTCACTCTGACCTTGAGAAAGCCAAAT 960
DB 1438 CAGCTGAGCATCAGAGGCAAAAGCAAACTCACTCTGACCTTGAGAAAGCCAAAT 1497
QY 961 GAACTGTCCCTGCTGCTCAGAGGTTGACAGAAAAGCAGAGCCCTCACTGCAACC 1020
DB 1498 GAACTGTCCCTGCTGCTCAGAGGTTGACAGAAAAGCAGAGCCCTCACTGCAACC 1557
QY 1021 TGTGCGACTGTGTAACCTCAGAGCAGACAGCAAAAGCCGTGATCCCGCAGCGTG 1080
DB 1558 TGTGCGACTGTGTAACCTCAGAGCAGACAGCAAAAGCCGTGATCCCGCAGCGTG 1617
QY 1081 CCCAGGTCCTCAGGCTGACAGCTGCTGCTTTAGAGACAGCCCGCTGTGACAGGGCTC 1140
DB 1618 CCCAGGTCCTCAGGCTGACAGCTGCTGCTTTAGAGAGACAGCCCGCTGTGACAGGGCTC 1677
QY 1141 AGTGGCTGACCTGTCCGACAGAGCTGAAAGCAGCAATTAAGCTCAAGCCTTCTTC 1200
DB 1678 AGTGGCTGACCTGTCCGACAGAGCTGAAAGCAGCAATTAAGCTCAAGCCTTCTTC 1737
QY 1201 TCTCTGATATCAATCACTTATATTAATGAGCCAGATGAGCATCTTACATGCTTC 1260
DB 1738 TCTCTGATATCAATCACTTATATTAATGAGCCAGATGAGCATCTTACATGCTTC 1797
QY 1261 TCTCTATCAGAAAGTCTTTGGAATCTACAAACCTTCCACTCTGTGATGGGACCAAC 1320
DB 1798 TCTCTATCAGAAAGTCTTTGGAATCTACAAACCTTCCACTCTGTGATGGGACCAAC 1857
QY 1321 AAGCTATGCAAGTTCTCCCTGTTGAGAACTATCGAGAGAGACTCCCGAAACCACTGCT 1380
DB 1858 AAGCTATGCAAGTTCTCCCTGTTGAGAACTATCGAGAGAGACTCCCGAAACCACTGCT 1917
QY 1381 GATTAAGAGAAAGCCAGCATCCCGAAGAAAGCTGACAGCCGACCTTACAGACCCAG 1440
DB 1918 GATTAAGAGAAAGCCAGCATCCCGAAGAAAGCTGACAGCCGACCTTACAGACCCAG 1977
QY 1441 AGCAAGCATTTGATTCGGTCAAGAACAGACAGATGAGCAACCGCCAGAGGTCCCTTTTA 1500
DB 1978 AGCAAGCATTTGATTCGGTCAAGAACAGACAGATGAGCAACCGCCAGAGGTCCCTTTTA 2037
QY 1501 TCTCCACTGATGAGAGTGGAGCTGAGAGCAATTAACACACAGCTTCTTTTCCGCG 1560

Db	2038	TCGCCATGTCATCGAAGTGGAGACGGGAGGACAAATTACCAACACAGCTTCTTTTGGC	2097
Qy	1561	CTTTCCACCAACCCAGCAGACCTTCACGAAATCTGCTGGCTTGGGCTTTAAGGCTGGAC	1620
Db	2098	CTTTCACACGACGACGACGACCTTCACGAAAGTCTGCTGGCTTGGGCTTTAAGGCTGGAC	2157
Qy	1621	TGGGAATATCTGGGCCCCCGAGACCTTACCCCTTCCCTGACCCAGCAGCTGGTATTTTGGC	1680
Db	2158	TGGGAATATCTGGGCCCCCGAGACCTTACCCCTTCCCTGACCCAGCAGCTGGTATTTTGGC	2217
Qy	1681	ACAGAGTCTTCACACTTCTACTCTGCGCTCAGGCATCTACCGAGGAGCGCAGTTAACTCT	1740
Db	2218	ACAGAGTCTTCACACTTCTACTCTGCGCTCAGGCATCTACCGAGGAGCGCAGTTAACTCT	2277
Qy	1741	GCTTACAGCTGACGCGACGCTGCGCACTTGCAGGAGACCAAGTCTATTCTGTGCGAGCGG	1800
Db	2278	GCTTACAGCTGACGCGACGCTGCGCACTTGCAGGAGACCAAGTCTATTCTGTGCGAGCGG	2337
Qy	1801	CAGAAAGCCAAAGTACAGAGCTGTACTCCGGGCGGAGCTGGCATGAAAGAGCCCTTTGAA	1860
Db	2338	CAGAAAGCCAAAGTACAGAGCTGTACTCCGGGCGGAGCTGGCATGAAAGAGCCCTTTGAA	2397
Qy	1861	AGCAGTTTAAACGAGAAAGCTGCCAAATGGAATTTGGAGAGACATCATGTACAGAAC	1920
Db	2398	AGCAGTTTAAACGAGAAAGCTGCCAAATGGAATTTGGAGAGACATCATGTACAGAAC	2457
Qy	1921	AGGTACCGGAAAGAGCTGGGGAAGTGGGAGCTGACGTACCTTTTGGGAGCATGAA	1980
Db	2458	AGGTACCGGAAAGAGCTGGGGAAGTGGGAGCTGACGTACCTTTTGGGAGCATGAA	2517
Qy	1981	ATCATTGAGGTCTCC	1995
Db	2518	ATCATTGAGGTCTCC	2532
RESULT 3			
LOCUS	AX260342	1998 bp	DNA
DEFINITION	Sequence 3 from Patent WO0173059.	linear	PAT 26-OCT-2001
ACCESSION	AX260342		
VERSION	AX260342.1	GI:16509305	
KEYWORDS			
SOURCE			
ORGANISM	Homo sapiens (human)		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
JOURNAL	1		
FEATURES	Meyers, R. A.		
source	38692 and 21117: dual specificity phosphatase molecules and uses therefor		
	Patent: WO 0173059-A 3 04-OCT-2001;		
	Millennium Pharmaceuticals, Inc. (US)		
	Location/Qualifiers		
	1..1998		
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
ORIGIN			
Query Match	94.9%;	Score 1893;	DB 6; Length 1998;
Best Local Similarity	99.9%;	Pred. No. 0;	
Matches 1993;	Conservative 0;	Mismatches 2;	Indels 0; Gaps 0;
Qy	1	ATGGCCCATGATGATGTGGAACCTCAATTTGTTACTGAGAGTGTGGTCTGTGGAA	60
Db	1	ATGGCCCATGATGATGTGGAACCTCAATTTGTTACTGAGAGTGTGGTCTGTGGAA	60
Qy	61	AGTGGAAACGAAAAAGTGTCTTAATGATAGCCGGCCATTTGTGGAATACATACATCC	120
Db	61	AGTGGAAACGAAAAAGTGTCTTAATGATAGCCGGCCATTTGTGGAATACATACATCC	120
Qy	121	CACATTTTGGAGGCATTAATATCAACTGCTCCAGCTTATGAAGGAAGTTGCAACG	180

Db	121	CAcATTtGGAAAGCCATTAATATCAACTGCTCCAGACTTATGAGAGGAAGGTTGCAACG	180
Qy	181	GACAAAGTGTAATTATCAGAGGCGATCCAGCAATTAGGCAAAATTAAGGTGACATGAT	240
Db	181	GACAAAGTGTATTTACAGAGCTATCCAGCAATTCAGCAAAATTAAGGTGACATGAT	240
Qy	241	TGCAGTCAGAAAGGTTGTAATTACGATCAAAAGCTCCCAAGATGTTGCCCTCTCTTTGA	300
Db	241	TGCAGTCAGAAAGGTTGTAATTACGATCAAAAGCTCCCAAGATGTTGCCCTCTCTTTGA	300
Qy	301	GACTGTTTTCTACGTACTTCTGGGTAACTGAGAGAAGCTTCACTCTGTTCACTG	360
Db	301	GACTGTTTTCTACGTACTTCTGGGTAACTGAGAGAAGCTTCACTCTGTTCACTG	360
Qy	361	CTTGCAGAGTGGGTTTGCTGAGTTCTCGTGTGTTTTCCCGGCGCTGCTGAGAGAAATCC	420
Db	361	CTTGCAGAGTGGGTTTGCTGAGTTCTCGTGTGTTTTCCCGGCGCTGCTGAGAGAAATCC	420
Qy	421	ACTTAGTCCCTACCTGCAATTTCTCAGCCTTGCTTACCTGTTGCCAACAATGGGCAAC	480
Db	421	ACTTAGTCCCTACCTGCAATTTCTCAGCCTTGCTTACCTGTTGCCAACAATGGGCAAC	480
Qy	481	CGAATTTCTTCCCAATCTTTATCTTGSGCTCCAGCGAGATGTCTCAACAAGAGCTGAT	540
Db	481	CGAATTTCTTCCCAATCTTTATCTTGSGCTCCAGCGAGATGTCTCAACAAGAGCTGAT	540
Qy	541	CAGCAGAAATGGAGTGGTTATGTGTTAATGCGCAGCTTACTGTGCCAAAGCTGACATT	600
Db	541	CAGCAGAAATGGAGTGGTTATGTGTTAATGCGCAGCAATACCTGTGCCAAAGCTGACATT	600
Qy	601	ATCCCCAGTCTCATTTCTCGCTGTGCTGTGAATGAACGTTTGTGAGAAATTTTG	660
Db	601	ATCCCCAGTCTCATTTCTCGCTGTGCTGTGAATGAACGTTTGTGAGAAATTTTG	660
Qy	661	CCGTGGTTGAGACAAATCACTAGATTTTCATTGAGAAAGCAAAAGCTCCATGATGTGTT	720
Db	661	CCGTGGTTGAGAAATCACTAGATTTTCATTGAGAAAGCAAAAGCTCCATGATGTGTT	720
Qy	721	CTAGTGAAGTGGTGGGATCTCCGCGTCCCGCACATCGATGCGCTACATCATG	780
Db	721	CTAGTGAAGTGGTGGGATCTCCGCGTCCCGCACATCGATGCGCTACATCATG	780
Qy	781	AAGAGGATGACATGCTCTTTAGATGAAGCTTACAGATTTGTGAAAGAAAAAGACTACT	840
Db	781	AAGAGGATGACATGCTCTTTAGATGAAGCTTACAGATTTGTGAAAGAAAAAGACTACT	840
Qy	841	ATATTCCTCAAACTTCAATTTTCTGGGCCAATCTCTGCACTATGAGAGAATTTAAGAC	900
Db	841	ATATTCCTCAAACTTCAATTTTCTGGGCCAATCTCTGCACTATGAGAGAATTTAAGAC	900
Qy	901	CAGACTGAGAGATCAGGGCCAAAGAGCAAACTCAAGCGTGCAGACTGGAAGACCAAT	960
Db	901	CAGACTGAGAGATCAGGGCCAAAGAGCAAACTCAAGCGTGCAGACTGGAAGACCAAT	960
Qy	961	GAACCTGTCCCTGCTGTCTCAGAGGATGACAGAAAAAGCGAGAGGCCCTTCACTCAACC	1020
Db	961	GAACCTGTCCCTGCTGTCTCAGAGGATGACAGAAAAAGCGAGAGGCCCTTCACTCAACC	1020
Qy	1021	TGTGCGCACTGTGTAACCTCAGAGGCGACGACAAAGGCCGTGCAATCCGCCACGGTG	1080
Db	1021	TGTGCGCACTGTGTAACCTCAGAGGCGACGACAAAGGCCGTGCAATCCGCCACGGTG	1080
Qy	1081	CCGAGCGGACCAGCGTGCAGCGCGTGCCTGTTAAGAGACAGCCCGCTGTATCAGGCGCTC	1140
Db	1081	CCGAGCGGACCAGCGTGCAGCGCGTGCCTGTTAAGAGACAGCCCGCTGTATCAGGCGCTC	1140
Qy	1141	AGTGGGCTGCACCTGTCCGAGACAGGCTGGAGAAGACATAAGCTCAAGCGTTCTTC	1200
Db	1141	AGTGGGCTGCACCTGTCCGAGACAGGCTGGAGAAGACATAAGCTCAAGCGTTCTTC	1200
Qy	1201	TCTCTGGAATCAAAATCAGTTTCAATATTCACGCCAGATGGCGAGATCTTTACATGAGCTTC	1260
Db	1201	TCTCTGGAATCAAAATCAGTTTCAATATTCAGCCAGATGGCGAGATCTTTACATGAGCTTC	1260

QY 1261 TCCTCATCAGAAATGCTTTGGAATACTAACAACCTTCCACTACTGTGATGGAGCAAC 1320
1261 TCCTCATCAGAAATGCTTTGGAATACTAACAACCTTCCACTACTGTGATGGAGCAAC 1320
QY 1321 AAGCTATGCAAGTTCTCCCTGTTCAAGAACTATCGAGAGACACTCCGAAACAGTCTT 1380
1321 AAGCTATGCAAGTTCTCCCTGTTCAAGAACTATCGAGAGACACTCCGAAACAGTCTT 1380
QY 1381 GATPAGAGAGAACCCAGATCCCAAGAACTGCAACCCGCGAGCTTTAGACAGCCAG 1440
1381 GATPAGAGAGAACCCAGATCCCAAGAACTGCAACCCGCGAGCTTTAGACAGCCAG 1440
QY 1441 AGCAAGGATTTGATTCGATCAAGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
1441 AGCAAGGATTTGATTCGATCAAGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
QY 1501 TCTCACTGATCGAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
1501 TCTCACTGATCGAAGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1560
QY 1561 CTTTCCACCAAGCAGAGACCTTCAAGAACTGTGCTGTGAGCTTGAAGGCTGTGAC 1620
1561 CTTTCCACCAAGCAGAGACCTTCAAGAACTGTGCTGTGAGCTTGAAGGCTGTGAC 1620
QY 1621 TCGGATATCTTGGCCCCCAGAGCTTACCCCTTCCCTGACAGAGAGAGAGAGAGAG 1680
1621 TCGGATATCTTGGCCCCCAGAGCTTACCCCTTCCCTGACAGAGAGAGAGAGAGAG 1680
QY 1681 ACAGAGTCTTCACTTCTTCTGCTTCAAGCACTTCAAGAGAGAGAGAGAGAGAGAG 1740
1681 ACAGAGTCTTCACTTCTTCTGCTTCAAGCACTTCAAGAGAGAGAGAGAGAGAGAG 1740
QY 1741 GCTTACAGCTGCAAGCAGCTGCCCCCTTGGAGAGAGAGAGAGAGAGAGAGAGAG 1800
1741 GCTTACAGCTGCAAGCAGCTGCCCCCTTGGAGAGAGAGAGAGAGAGAGAGAGAG 1800
QY 1801 CAGAGAGCAG 1860
1801 CAGAGAGCAG 1860
QY 1861 AAGCAGTTTAAACGAG 1920
1861 AAGCAGTTTAAACGAG 1920
QY 1921 AGGTCAAGGAG 1980
1921 AGGTCAAGGAG 1980
QY 1981 ATCATGAGGCTTCC 1995
1981 ATCATGAGGCTTCC 1995
Db 1981 ATCATGAGGCTTCC 1995

RESULT 4
AX278461 3059 bp DNA linear PAT 02-NOV-2001
LOCUS AX278461 Sequence 1 from Patent WO0177340.
DEFINITION AX278461
ACCESSION AX278461.1 GI:16605915
VERSION
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE
1 Duecker, K.
AUTHORS Identification of a dual specificity phosphatase: dusp-10
TITLE Patent: WO 0177340-A, 18-Oct-2001;
JOURNAL MERCK PATENT GmbH (DE)
FEATURES
1. 3059
Location/Qualifiers
source /organism="Homo sapiens"

/mol_type="unassigned DNA"
/db_xref="taxon:9606"
127..2124
/note="unassigned protein product"

CDS

/codon_start=1
/protein_id="CAD10545.1"
/db_xref="GI:16605916"
/db_xref="REMBL:CAD10545"
/translation="MAHEMIGQIVTERVALLSEGTGKVLIDSRPVEVNTSHLE
AININCSKMKRRLQODKRLITELIOLHSAKHVYDIDSGKVVYVYDSSQDVASLSDC
PLTVLLGKLEKSFNSVHLAAGPAEPRCPGLCEKSLVLPCLISQCLPLANISPT
RLPLNLYLCCQDVLNKELMQNGIGVYLANASNTCEKPDIPESHFLRPVNDSPKX
ILPMLDKSVDFIEKASNGCVLHCLAGISRSATIALIYIMKRMDSIDEXYR FYKE
KPTTSPNPNFGLDYEKIKNQTGAGSPSKLHLLEKPNBPVAVSEGGQKSE
TLPSPCADSATSEAGORPVHAPVSPVOPSLLEPSLIVOLISGLHLSADRLD
SKLRKSRSLDKSYISASMAASHGSSSDALEYKRPSTTLDTNKLCPSPVQZ
LSEQTPETSPDKEASIPKRLQTPARPSDSQRLHSVRSSTSGTQRLSLSPATSEHS
VEDNHTSFLFGLSTSOHLTASAGLGMWSHDLAPGTSTPSLTSVFPATSESHF
YSASAIYGGASVSAVSCQLPTCGQVSVRRQKPSDRADSRSRSMHESEPFKEQFK
RSCQMEFSESIMSENRSEBELGVQSSSFGSMELIIEVS"

ORIGIN

Query Match 94.9%; Score 1893; DB 6; Length 3059;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGCCCCATGAGATGATTTGGAATCTCAAAATTTGTAATCTGAGAGAGTGGTGGCTCTGCTGAA 60
127 ATGCCCCATGAGATGATTTGGAATCTCAAAATTTGTAATCTGAGAGAGTGGTGGCTCTGCTGAA 186
QY 61 AGTGAACGGAAGAAAGTGTCTGTAATTTGATAGCCGSCATTTGTGAATATACATATCC 120
187 AGTGAACGGAAGAAAGTGTCTGTAATTTGATAGCCGSCATTTGTGAATATACATATCC 246
QY 121 CACATTTTGAAGCCATTAATATCAACCTCTCAAGCTTATGAGAGGAGTTGCAACAG 180
247 CACATTTTGAAGCCATTAATATCAACCTCTCAAGCTTATGAGAGGAGTTGCAACAG 306
QY 181 GACAAAGTGTAAATTAACAGAGCTCATCCAGCACTTCCAGCAATATGAGTTGATATGAT 240
307 GACAAAGTGTAAATTAACAGAGCTCATCCAGCACTTCCAGCAATATGAGTTGATATGAT 366
QY 241 TGCAGTCAAGAGTGTATGTTTACATCAAACTCCCAAGATGTTGCTCTCTCTTCA 300
367 TGCAGTCAAGAGTGTATGTTTACATCAAACTCCCAAGATGTTGCTCTCTCTTCA 426
QY 301 GACTGTTTCTCACTGTACTTCTGGGTAAACTGAGAGAGAGCTTCAACTCTGTTCACTG 360
427 GACTGTTTCTCACTGTACTTCTGGGTAAACTGAGAGAGAGCTTCAACTCTGTTCACTG 486
QY 361 CTTCAGAGTGGGTTTGTCTAGATTTCTCGTTGTTTCCCTGAGCTCTGTGAAGAAATTC 420
487 CTTCAGAGTGGGTTTGTCTAGATTTCTCGTTGTTTCCCTGAGCTCTGTGAAGAAATTC 546
QY 421 ACTCATGCTCCTACCTGCACTTCTCAGCTTGTCTTACCTGTTGCCAAATTTGGGCCAAC 480
547 ACTCATGCTCCTACCTGCACTTCTCAGCTTGTCTTACCTGTTGCCAAATTTGGGCCAAC 606
QY 481 CGAATTTCTCCCAATCTTATCTTGGCTGCCAGCAGAGATGCTCTCAACAGAGAGCTGATA 540
607 CGAATTTCTCCCAATCTTATCTTGGCTGCCAGCAGAGATGCTCTCAACAGAGAGCTGATA 666
QY 541 CAGCAGATGAGATGTTATGTTTAAATGCGAGATATCTGTGCCAAAGCTGACTTT 600
667 CAGCAGATGAGATGTTATGTTTAAATGCGAGATATCTGTGCCAAAGCTGACTTT 726
QY 601 ATCCCGAGTCTCATTTCTGCGTGGCTGCTGGAATGACAGCTTTTGTGAGAAATTTTG 660
727 ATCCCGAGTCTCATTTCTGCGTGGCTGCTGGAATGACAGCTTTTGTGAGAAATTTTG 786
QY 661 CCGTGGTTGACAAATCAGTAGATTTCAATGAGAAAGCAAAAGCTCCAAATGATGTGTT 720
787 CCGTGGTTGACAAATCAGTAGATTTCAATGAGAAAGCAAAAGCTCCAAATGATGTGTT 846

QY 721 CTATGCACTGTTAGTGGGATCTCCGCTCCGSCACATCGCTATCGCTCATCATG 780
DB 847 CTATGCACTGTTAGTGGGATCTCCGCTCCGSCACATCGCTATCGCTCATCATG 906
QY 781 AAGAGATGCACTGCTCTTATGATGAAGCTTACAGATTTGTGAAGAAAAAGACTACT 840
DB 907 AAGAGATGCACTGCTCTTATGATGAAGCTTACAGATTTGTGAAGAAAAAGACTACT 966
QY 841 ATATCTCAAACTTCAATTTTCTGGGCAACTCTGTGACATATGAGAAAGATTAAAGAC 900
DB 967 ATATCTCAAACTTCAATTTTCTGGGCAACTCTGTGACATATGAGAAAGATTAAAGAC 1026
QY 901 CAGACTGAGCATCAGGGCCAAAGAGCAAACTCAGCTGTGACCTGAGAGCAAAAT 960
DB 1027 CAGACTGAGCATCAGGGCCAAAGAGCAAACTCAGCTGTGACCTGAGAGCAAAAT 1086
QY 961 GAACCTGTCCCTGCTGTCTCAGAGGCTGGA CAGAAAAAGCAGACGCCCCCTCACTCCACC 1020
DB 1087 GAACCTGTCCCTGCTGTCTCAGAGGCTGGA CAGAAAAAGCAGACGCCCCCTCACTCCACC 1146
QY 1021 TGTGCGGACTCTGCTACCTCAGAGGCGAGCAAAAGCCGCTGATCCCCGACGCTG 1080
DB 1147 TGTGCGGACTCTGCTACCTCAGAGGCGAGCAAAAGCCGCTGATCCCCGACGCTG 1206
QY 1081 CCCAGCGTGCACGCGTGCAGCGCTGCTGTTAGAGGACAGCCCGCTGTTAGAGGCTTC 1140
DB 1207 CCCAGCGTGCACGCGTGCAGCGCTGCTGTTAGAGGACAGCCCGCTGTTAGAGGCTTC 1266
QY 1141 AGTGGGCTGCACCTGTCCGAGACAGGCTGGAAGACAGCAATTAAGCTCAAGCGTTCTTC 1200
DB 1267 AGTGGGCTGCACCTGTCCGAGACAGGCTGGAAGACAGCAATTAAGCTCAAGCGTTCTTC 1326
QY 1201 TCTCTGGATTCATCAATCAGTTTATATTCATTCAGCAGAGATGGACATCTTACATGGCTTC 1260
DB 1327 TCTCTGGATTCATCAATCAGTTTATATTCATTCAGCAGAGATGGACATCTTACATGGCTTC 1386
QY 1261 TCTCTCATCAGAAAGATCTTTGGAATACTCAAACTTCCACTACTCTGATGGAGCAAC 1320
DB 1387 TCTCTCATCAGAAAGATCTTTGGAATACTCAAACTTCCACTACTCTGATGGAGCAAC 1446
QY 1321 AAGCTATGCCAGTTCCTCCCTGTTCCAGAACTATCGGAGCAGACTCCGAAACCACTCT 1380
DB 1447 AAGCTATGCCAGTTCCTCCCTGTTCCAGAACTATCGGAGCAGACTCCGAAACCACTCT 1506
QY 1381 GATTAAGAGAGAGCAGCATCCCAAGAAAGCTGCAGACCGGCTTCAACAGCGCAG 1440
DB 1507 GATTAAGAGAGAGCAGCATCCCAAGAAAGCTGCAGACCGGCTTCAACAGCGCAG 1566
QY 1441 AGCAAGCAGTTGCAATTCCGTCAGAAACCAAGCAGAGTGCACCGCCAGAGTCCCTTTTA 1500
DB 1567 AGCAAGCAGTTGCAATTCCGTCAGAAACCAAGCAGAGTGCACCGCCAGAGTCCCTTTTA 1626
QY 1501 TCTCCACTGATGGAAGTGGAGCGTGGAGACATTAACAACAAGCTTCTTTTCGCG 1560
DB 1627 TCTCCACTGATGGAAGTGGAGCGTGGAGACATTAACAACAAGCTTCTTTTCGCG 1686
QY 1561 CTCTTCACGACGACGACGACCTCAAGAGTCAGCTGCGCTGGGCTTAAAGGGCTGGAC 1620
DB 1687 CTCTTCACGACGACGACGACCTCAAGAGTCAGCTGCGCTGGGCTTAAAGGGCTGGAC 1746
QY 1621 TCGGATATCTGGCCCCCAGACCTTCAACCTTCCCTGACCAAGAGCTGATTTTTCGC 1680
DB 1747 TCGGATATCTGGCCCCCAGACCTTCAACCTTCCCTGACCAAGAGCTGATTTTTCGC 1806
QY 1681 AAGAGTCTTCACTTCTGCTGCTGACCATCTAGAGGAGAGGAGCTGATCTCT 1740
DB 1807 AAGAGTCTTCACTTCTGCTGCTGACCATCTAGAGGAGAGGAGCTGATCTCT 1866
QY 1741 GCCTACAGCTGAGCAGCTGCCCACTTGGAGAGCAAGTATTTCTGTGGCAGGAGG 1800
DB 1867 GCCTACAGCTGAGCAGCTGCCCACTTGGAGAGCAAGTATTTCTGTGGCAGGAGG 1926

QY 1801 CAGAGCCAAAGTGAACAGAGCTGACTCGGCGGAGAGCTGGCATGAAGAGAGCCCTTTGA 1860
DB 1927 CAGAGCCAAAGTGAACAGAGCTGACTCGGCGGAGAGAGCTGGCATGAAGAGAGCCCTTTGA 1986
QY 1861 AAGCATTTTAAACGAGAGAGCTGCCAAATGGAATTTGAGAGAGCATGTCAGAGAAC 1920
DB 1987 AAGCATTTTAAACGAGAGAGCTGCCAAATGGAATTTGAGAGAGCATGTCAGAGAAC 2046
QY 1921 AGGTACCGGAAAGACTGGGAAAGTGGGCACTGACTGATCTTTTCCGGCAGCATGGA 1980
DB 2047 AGGTACCGGAAAGACTGGGAAAGTGGGCACTGACTGATCTTTTCCGGCAGCATGGA 2106
QY 1981 ATCATTGAGGTCTCC 1995
DB 2107 ATCATTGAGGTCTCC 2121

RESULT 5
AX441210 3496 bp DNA linear PAT 28-JUN-2002
LOCUS AX441210
DEFINITION Sequence 1 from Patent WO0226997.
ACCESSION AX441210
VERSION AX441210.1 GI:2165766
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1
Luche, R.M. and Wei, B.
Dsp-16 dual-specificity phosphatase
Patent: WO 0226997-A 1 04-APR-2002;
JOURNAL
Ceptyr, Inc. (us)

FEATURES
source
1..3496
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Query Match 94.9%; Score 1893; DB 6; Length 3496;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGGCCCATGAGATTTGAATCAATTTGTTACTGAGAGTGTGGCTCTGCTGGA 60
DB 562 ATGGCCCATGAGATTTGAATCAATTTGTTACTGAGAGTGTGGCTCTGCTGGA 621
QY 61 AGTGAACGGAAGAGTGTGCTAATTGATGACCGGCATTTGTGAATCAATCAATCC 120
DB 622 AGTGAACGGAAGAGTGTGCTAATTGATGACCGGCATTTGTGAATCAATCAATCC 681
QY 121 CACATTTTGAAGCATTAATATCACTGCTCAAGCTTATGAAGCAGAGTTGCAAG 180
DB 682 CACATTTTGAAGCATTAATATCACTGCTCAAGCTTATGAAGCAGAGTTGCAAG 741
QY 181 GACAAAGTTAATTAAGAGCTCATCAGATTCACAGAAACATAAGTTGACATTGAT 240
DB 742 GACAAAGTTAATTAAGAGCTCATCAGATTCACAGAAACATAAGTTGACATTGAT 801
QY 241 TGCAGTGAAGAGTTGATGTTAGATCAAGAGCTCCCAAGATGTGGCTCTCTCTTCA 300
DB 802 TGCAGTGAAGAGTTGATGTTAGATCAAGAGCTCCCAAGATGTGGCTCTCTCTTCA 861
QY 301 GACTGTTTTCTCACTGTAATTTGAGTAACTGAGAAAGCTTCAACTGTTCACCTG 360
DB 862 GACTGTTTTCTCACTGTAATTTGAGTAACTGAGAAAGCTTCAACTGTTCACCTG 921
QY 361 CTGCAAGTGGGTTTGTGAGATTTCTCGTGTGTTCCCTGGGCTCTGTAAGGAAATCC 420
DB 922 CTGCAAGTGGGTTTGTGAGATTTCTCGTGTGTTCCCTGGGCTCTGTAAGGAAATCC 981
QY 421 ACTCTAGTCCCTTCACTGCAATTTCTCAGCCTTGCTTACCTGTTCCAATTTGGCCAAC 480

Db 982 ACCTAGTCCCTACCTGATTTCTCAGCCCTTGCTACCTGTGCAAACTTGGGCCAAC 1041
 Qy 481 CGAATCTTCCCAATCTTTATCTTGGCTGCCAGCAGATGTCTCTCAACAGAGCTGATA 540
 Db 1042 CGAATCTTCCCAATCTTTATCTTGGCTGCCAGCAGATGTCTCTCAACAGAGCTGATA 1101
 Qy 541 CAGCAGAAATGGATGGTATGTGTAAATGCCAGCTATACCTGTCCAAAGCCGTGACTT 600
 Db 1102 CAGCAGAAATGGATGGTATGTGTAAATGCCAGCTATACCTGTCCAAAGCCGTGACTT 1161
 Qy 601 ATCCCGAGTCTGATTTCTGCGTGTGCTGTGATGACAGCTTTGTGAAAATTTTG 660
 Db 1162 ATCCCGAGTCTGATTTCTGCGTGTGCTGTGATGACAGCTTTGTGAAAATTTTG 1221
 Qy 661 CGGTGTTGGAACAATAGTAGATTTTATGAGAAAGCAAAAGCTTCAATGATGTGT 720
 Db 1222 CGGTGTTGGAACAATAGTAGATTTTATGAGAAAGCAAAAGCTTCAATGATGTGT 1281
 Qy 721 CTAGTGACTGTTAGCTGGGATCTCCGCTCCGCCACATCGGTATCGCTATCATCATG 780
 Db 1282 CTAGTGACTGTTAGCTGGGATCTCCGCTCCGCCACATCGGTATCGCTATCATCATG 1341
 Qy 781 AAGAGATGACATGTCTTTAGATGAGCTTAAAGATTTGTGAAAAGAAAAGACCTACT 840
 Db 1342 AAGAGATGACATGTCTTTAGATGAGCTTAAAGATTTGTGAAAAGAAAAGACCTACT 1401
 Qy 841 ATATCTCCAACTTCAATTTTCTGGGCAACTCTCGGACTATGAGAAAGATTAAGAAC 900
 Db 1402 ATATCTCCAACTTCAATTTTCTGGGCAACTCTCGGACTATGAGAAAGATTAAGAAC 1461
 Qy 901 CAGACTGAGACATCAGGGCCAAAGAGAACTAAGCTGTGACCTGTGAGAGCCCAAT 960
 Db 1462 CAGACTGAGACATCAGGGCCAAAGAGAACTAAGCTGTGACCTGTGAGAGCCCAAT 1521
 Qy 961 GAACTGTCCCTGCTGTCTCAGAGGCTGACAGAAAAGCAGAGCCCTCAGTCCACC 1020
 Db 1522 GAACTGTCCCTGCTGTCTCAGAGGCTGACAGAAAAGCAGAGCCCTCAGTCCACC 1080
 Qy 1021 TGTGCGACTCTGTCTCAGAGGCTGACAGAAAAGCAGAGCCCTCAGTCCACC 1080
 Db 1582 TGTGCGACTCTGTCTCAGAGGCTGACAGAAAAGCAGAGCCCTCAGTCCACC 1641
 Qy 1081 CCCAGGCTGCCAGCCGTGACAGGCTGTGAGAGACCCCGCTGTGACAGGCGCTC 1140
 Db 1642 CCCAGGCTGCCAGCCGTGACAGGCTGTGAGAGACCCCGCTGTGACAGGCGCTC 1701
 Qy 1141 AGTGGCTGCACTGTCCGAGACAGGCTGAGAAAGCAATAGCTCAAGCCTTCTTC 1200
 Db 1702 AGTGGCTGCACTGTCCGAGACAGGCTGAGAAAGCAATAGCTCAAGCCTTCTTC 1761
 Qy 1201 TCTCTGATATCAATAGTTTATATTCAGCCAGCAGTGCAGATCTTCAATGCTTC 1260
 Db 1762 TCTCTGATATCAATAGTTTATATTCAGCCAGCAGTGCAGATCTTCAATGCTTC 1821
 Qy 1261 TCTCTCATCAGAAATGTTTGAATATCAAACTTCCACTACTGTGATGGAGCCAC 1320
 Db 1822 TCTCTCATCAGAAATGTTTGAATATCAAACTTCCACTACTGTGATGGAGCCAC 1380
 Qy 1321 AAGCTATGCAAGTTCTCCCTGTTACAGAACTATCGAGAGAGCTCCCGAAACAGTCC 1380
 Db 1882 AAGCTATGCAAGTTCTCCCTGTTACAGAACTATCGAGAGAGCTCCCGAAACAGTCC 1941
 Qy 1381 GATTAAGAGAGACCCAGCATCCCAAGAGCTGACAGCCGAGCCTTACAGACCCAG 1440
 Db 1942 GATTAAGAGAGACCCAGCATCCCAAGAGCTGACAGCCGAGCCTTACAGACCCAG 2001
 Qy 1441 AGCAAGGATTTGATTTGGTTCAGAAACAGCAGAGCTGACAGCCGAGCCTTACAG 1500
 Db 2002 AGCAAGGATTTGATTTGGTTCAGAAACAGCAGAGCTGACAGCCGAGCCTTACAG 2061
 Qy 1501 TCTCACTGATGAGTGGAGGCTGAGAGCAATTAACACACAGCTTCTTTTCGCG 1560

Db 2062 TCTCACTGATCAGTGAAGTGGAGCGTGGAGAGCAATTAACACACAGCTTCTTTTCGCG 2121
 Qy 1561 CTTTCACAGACGAGAGACCTCAGAGTGTGCTGGGCTTAAAGGCTGGAC 1620
 Db 2122 CTTTCACAGACGAGAGACCTCAGAGTGTGCTGGGCTTAAAGGCTGGAC 2181
 Qy 1621 TCGATATCTTGGCCCCCAGACCTTACCCCTTCCCTGACACAGAGCTGTATTTTGCC 1680
 Db 2182 TCGATATCTTGGCCCCCAGACCTTACCCCTTCCCTGACACAGAGCTGTATTTTGCC 2241
 Qy 1681 ACAGAGCTTCACTCTTCTGCTGCTGACAGCCATTTACGAGAGAGCTGCACTACT 1740
 Db 2242 ACAGAGCTTCACTCTTCTGCTGCTGACAGCCATTTACGAGAGAGCTGCACTACT 2301
 Qy 1741 GCCTAAGCTGACAGCCAGCTGCCCCATTTGCGAGAGCAAGTCTATTTGTGCGCAGCGG 1800
 Db 2302 GCCTAAGCTGACAGCCAGCTGCCCCATTTGCGAGAGCAAGTCTATTTGTGCGCAGCGG 2361
 Qy 1801 CAGAACCCAAAGTACAGAGCTGACTGCGCGGAGCTGGCATGAGAGAGCCCTTTGAA 1860
 Db 2362 CAGAACCCAAAGTACAGAGCTGACTGCGCGGAGCTGGCATGAGAGAGCCCTTTGAA 2421
 Qy 1861 AAGCAGTTTAAAGCAGAACTGCAAAATGGAATTTGAGAGAGCATATGTCAGAAAC 1920
 Db 2422 AAGCAGTTTAAAGCAGAACTGCAAAATGGAATTTGAGAGAGCATATGTCAGAAAC 2481
 Qy 1921 AGGTACGGGAAAGAGCTGGGAAAGTGGGAGTCACTTACCTTTTCCGGCAGCATGAA 1980
 Db 2482 AGGTACGGGAAAGAGCTGGGAAAGTGGGAGTCACTTACCTTTTCCGGCAGCATGAA 2541
 Qy 1981 ATCATGAGTCTCC 1995
 Db 2542 ATCATGAGTCTCC 2556

RESULT 6
 AB052156
 LOCUS Homo sapiens MKP-7 mRNA for MAPK phosphatase-7, complete cds.
 DEFINITION
 AB052156
 VERSION
 AB052156.1 GI:13548676
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 AUTHORS
 Masuda, K., Shima, H., Watanabe, M. and Kikuchi, K.
 TITLE
 MKP-7, a novel mitogen-activated protein kinase phosphatase,
 functions as a shuttle protein
 JOURNAL
 J. Biol. Chem. 276 (42), 39002-39011 (2001)
 MEDLINE
 21486429
 PUBMED
 11489891
 REFERENCE
 2 (bases 1 to 3521)
 AUTHORS
 Masuda, K., Shima, H. and Kikuchi, K.
 TITLE
 Direct Substitution
 Submitted (05-DEC-2000) Kouhei Masuda, Institute for Genetic
 Medicine, Division of Biochemical Oncology and Immunology, Kita-ku
 Kitais-Jou 1-17-1, Sapporo, Hokkaido 060-0815, Japan
 (E-mail: kouhei.masuda@kitai.ac.jp, Tel: 81-11-706-5536,
 Fax: 81-11-707-6839)
 FEATURES
 source
 1..3521
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="12"
 /cell_line="Jurkat"
 1..3521
 /gene="MKP-7"
 564..2561
 /gene="MKP-7"
 /codon_start=1


```
/product="MAPK phosphatase-7"  
/protein_id="BAB40814.1"  
/db_xref="GI:13548677"  
/translation="MAHEMIGDIYTERLVALLIESTERYLLIDSPFVEYINTSHILE  
AININCKLMKRRLQODKYLITELIGHSAKHVIDIDSKVWVYDSSQDVASLSDC  
FLVLLGKLEKSFNSVHLIAGFAEFPGLCEKSGFLVPTCLSPQCLPVANIGPT  
RIIPNLYIGCORLVNKEIMQONGIGVLANSTCEKPIFESHRLSPVNDVDFEKE  
ILEPMLKSYDFIEKAKASNGCVLHCLAGISRSATIAIYIKRMMSLIDEARFYKE  
KRPTISPNPFLGOLLDYREKLIKNONGASGPKSLILHLIKENPEVPAIVSGGQCE  
TPISPCPADSATIEAAGQRPVHPASVPSPVSPDILLESPLVQALSGHLSADRLSD  
SNLKRFSFLDIKVSYSASMAASLHGFSSBEDALRYKPSITLIDTNKLCOFSPQVE  
LEOPEPESPDKEBASIPKLIQARPSDSQSKRLHSVRTSSGRTSLRSLSPVDS  
VEDNHTSFLFGLSTISQHLTKSAGLKGMSHDIAPQTSSTLSYSYFATSESHF  
YSASAIYGSASAYSAYSCGLPTCPDOVSVRRORPSRADRSRMHESPEKQFK  
RRCOMEPEBSIMSENRSRBELCKVSGSFSGSMETIIVS"
```

ORIGIN

```
Query Match      94.9%; Score 1893; DB 9; Length 3521;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      1 ATGGCCCATGAGTGAATGGAATCAATTTGTTACTGAGAGTTGGTGGCTCTGCTGGAA 60  
DB      564 ATGGCCCATGAGTGAATGGAATCAATTTGTTACTGAGAGTTGGTGGCTCTGCTGGAA 623  
QY      61 AGTGAACGGAAAAAGTGTCTAATTGATGAGCGGCGCAATTTGGGAATPACAATACATCC 120  
DB      624 AGTGAACGGAAAAAGTGTCTAATTGATGAGCGGCGCAATTTGGGAATPACAATACATCC 683  
QY      121 CACATTTTGGAAACCAATTAATCAATGCTCCAGCTTAAGAAGCAAGTTGCAACAG 180  
DB      684 CACATTTTGGAAACCAATTAATCAATGCTCCAGCTTAAGAAGCAAGTTGCAACAG 743  
QY      181 GACAAAGTGTATTATACAGAGCTCATCCAGATTGAGCGAAACATTAAGTTGACATTGAT 240  
DB      744 GACAAAGTGTATTATACAGAGCTCATCCAGATTGAGCGAAACATTAAGTTGACATTGAT 803  
QY      241 TGCAGTCAGAAAGTTGATGATGATCAAAAGCTCCCAAGATGTTGCTCTCTCTTCA 300  
DB      804 TGCAGTCAGAAAGTTGATGATGATCAAAAGCTCCCAAGATGTTGCTCTCTCTCTTCA 863  
QY      301 GACTGTTTCTCACTGACTTCTGGGTAAACTGGAGAAGAGCTTCAACTCTGTTCACTG 360  
DB      864 GACTGTTTCTCACTGACTTCTGGGTAAACTGGAGAAGAGCTTCAACTCTGTTCACTG 923  
QY      361 CTTCGACGTTGGTTTGTGAGTTCTCTGTTGTTTCCCTGGCTCTGTGAGAGAAATCC 420  
DB      924 CTTCGACGTTGGTTTGTGAGTTCTCTGTTGTTTCCCTGGCTCTGTGAGAGAAATCC 983  
QY      421 ACTCTAGTCCCTACCTGCAATTTCTCAGCCCTTGCTTACCTGTTGCGAAACATTTGGGCGCAAC 480  
DB      984 ACTCTAGTCCCTACCTGCAATTTCTCAGCCCTTGCTTACCTGTTGCGAAACATTTGGGCGCAAC 1043  
QY      481 CGAATTCTTCCCAATCTTTATCTTGGCTGCGACAGAGATGTCCTCAACAAGAGCTGATAT 540  
DB      1044 CGAATTCTTCCCAATCTTTATCTTGGCTGCGACAGAGATGTCCTCAACAAGAGCTGATAT 1103  
QY      541 CAGCAGAAATGGGATGTTATGTTTAAATGCCAGCTATACCTGTCCAAAGCCTGACTTT 600  
DB      1104 CAGCAGAAATGGGATGTTATGTTTAAATGCCAGCAATATACCTGTCCAAAGCCTGACTTT 1163  
QY      601 ATCCCGAGATCTAATTTCCGCGTGGCTGCGTGAATGAGCAGCTTTGTGAGAAATTTTG 660  
DB      1164 ATCCCGAGATCTAATTTCCGCGTGGCTGCGTGAATGAGCAGCTTTGTGAGAAATTTTG 1223  
QY      661 CCGTGGTTGGACAAATCAGTAGATTTCATTGAGAAAGCAAAAGCCTCAATGGATGTTT 720  
DB      1224 CCGTGGTTGGACAAATCAGTAGATTTCATTGAGAAAGCAAAAGCCTCAATGGATGTTT 1283  
QY      721 CTATGTCAGCTGTTTAACTGGGATCTCCCGCTCCGCAACATGCTATCGCTTACATCATG 780  
DB      1284 CTATGTCAGCTGTTTAACTGGGATCTCCCGCTCCGCAACATGCTATCGCTTACATCATG 1343
```

```
QY      781 AAGAGATGAGCATGCTTTAGATGAGTCAAGATTGTTGAAGAAAAAGCACTACT 840  
DB      1344 AAGAGATGAGCATGCTTTAGATGAGTCAAGATTGTTGAAGAAAAAGCACTACT 1403  
QY      841 ATATCTCCAACTTCAATTTTCTGGGCCCAACTCTCTGACATATGAGAGAGATTAAAGC 900  
DB      1404 ATATCTCCAACTTCAATTTTCTGGGCCCAACTCTCTGACATATGAGAGAGATTAAAGC 1463  
QY      901 CAGACTGGAGCTCAAGGGCGCAAGAGCAAACTCAAGCTGCTGACCTGAGAGAGCAAT 960  
DB      1464 CAGACTGGAGCTCAAGGGCGCAAGAGCAAACTCAAGCTGCTGACCTGAGAGAGCAAT 1523  
QY      961 GAACCTGTCTCCCTGCTCTCAAGAGGTGAGCAGAAAAGCAGACCCCTCAGTCCACCC 1020  
DB      1524 GAACCTGTCTCCCTGCTCTCAAGAGGTGAGCAGAAAAGCAGACCCCTCAGTCCACCC 1583  
QY      1021 TGTGCCGACTGTGCTACTCTCAGAGGCAAGAGCAAAAGCCGTGCATCTCCGCAAGCTG 1080  
DB      1584 TGTGCCGACTGTGCTACTCTCAGAGGCAAGAGCAAAAGCCGTGCATCTCCGCAAGCTG 1643  
QY      1081 CCCAGCGGCCCGAGGGTGCAGCGCGTGGCTTTAGAGACAGCGCGCTGGTACAGCGGCTC 1140  
DB      1644 CCCAGCGGCCCGAGGGTGCAGCGCGTGGCTTTAGAGACAGCGCGCTGGTACAGCGGCTC 1703  
QY      1141 AGTGGGCTGCACCTGTCCGACAGACAGGCTGGAAGA CAGCAATTAAGCTCAAGCGTTCTTTC 1200  
DB      1704 AGTGGGCTGCACCTGTCCGACAGACAGGCTGGAAGA CAGCAATTAAGCTCAAGCGTTCTTTC 1763  
QY      1201 TCTCTGATATCAATCAGTTTCATATTCAGCCAGCATGAGCAGCATCTTACATGCGCTTC 1260  
DB      1764 TCTCTGATATCAATCAGTTTCATATTCAGCCAGCATGAGCAGCATCTTACATGCGCTTC 1823  
QY      1261 TCCCTGATGAGAGAGCTTTGGGAATCTAACAACTTCCACTGATGGAATGGGACCAAC 1320  
DB      1824 TCCCTGATGAGAGAGCTTTGGGAATCTAACAACTTCCACTGATGGAATGGGACCAAC 1883  
QY      1321 AAGCTATGCAAGTTTCTCCCTGTTGAGGAACATATCGAGACAGACTCCGAAACAGTCT 1380  
DB      1884 AAGCTATGCAAGTTTCTCCCTGTTGAGGAACATATCGAGACAGACTCCGAAACAGTCT 1943  
QY      1381 GATTAAGAGAGAGCAGACATCCCAAGAGCTGCAAGCCGCGCAAGCTTCAAGACAGCAG 1440  
DB      1944 GATTAAGAGAGAGCAGACATCCCAAGAGCTGCAAGCCGCGCAAGCTTCAAGACAGCAG 2003  
QY      1441 AGCAAGCAATTCATTCGTCAGAACCAAGAGCAATGAGCAACGCGCAAGAGCTCTTTTA 1500  
DB      2004 AGCAAGCAATTCATTCGTCAGAACCAAGAGCAATGAGCAACGCGCAAGAGCTCTTTTA 2063  
QY      1501 TCTCCAGTCGATCGAAGTGGAGCGTGGAGGCAATTAACAACAGCTTCTTTTGGGC 1560  
DB      2064 TCTCCAGTCGATCGAAGTGGAGCGTGGAGGCAATTAACAACAGCTTCTTTTGGGC 2123  
QY      1561 CTTTCCACCAAGCAGCAGACCTCAAGAGTGTGCTGCGCTTGAGGCTTAAAGGCTGGCAC 1620  
DB      2124 CTTTCCACCAAGCAGCAGACCTCAAGAGTGTGCTGCGCTTGAGGCTTAAAGGCTGGCAC 2183  
QY      1621 TGGGATATCTGGCCCCCCCCAAGACTCTTACCCCTTCCGACCAAGAGAGCTGATTTTGGC 1680  
DB      2184 TGGGATATCTGGCCCCCCCCAAGACTCTTACCCCTTCCGACCAAGAGAGCTGATTTTGGC 2243  
QY      1681 ACAGAGTCTTCACTTCACTCTGCTCAGGCATCTTACGAGGAGAGTGGCCAGTTACTCT 1740  
DB      2244 ACAGAGTCTTCACTTCACTCTGCTCAGGCATCTTACGAGGAGAGTGGCCAGTTACTCT 2303  
QY      1741 GCTTACAGCTGAGCAGCTGCGCACTTTCGAGAGCAAGTCTATTCTGTGCGCAGGCGG 1800  
DB      2304 GCTTACAGCTGAGCAGCTGCGCACTTTCGAGAGCAAGTCTATTCTGTGCGCAGGCGG 2363  
QY      1801 CAGAAAGCAAGTGAAGAGCTGACTCGCGGCGGAGAGCTGGCAGTGAAGAGAGCCCTTTGAA 1860  
DB      2364 CAGAAAGCAAGTGAAGAGCTGACTCGCGGCGGAGAGCTGGCAGTGAAGAGAGCCCTTTGAA 2423  
QY      1861 AAGCAGTTTAAACGCAAGAGCTGCCAATTGAAATTTGAGAGAGATCATGTCAGAGAAC 1920
```

Db 2424 AACAGCTTTAAAGCAGAGAGCTCCAAATGGAATTTGAGAGAGCATCATGTCAGAGAAC 2483
QY 1921 AGGTACCGGGAAGAGCTGGGGAAGTGGGAGTCACTGCTTTTGGGACAGATGAA 1980
Db 2484 AGGTACCGGGAAGAGCTGGGGAAGTGGGAGTCACTGCTTTTGGGACAGATGAA 2543
QY 1981 ATGATTGAGGTCTCC 1995
Db 2544 ATCATTGAGGTCTCC 2558

RESULT 7
AX260340 3544 bp DNA linear PAT 26-OCT-2001
DEFINITION Sequence 1 from Patent WO0173059.
ACCESSION AX260340
VERSION AX260340.1 GI:16509303
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1
AUTHORS Meyers, R.A.
TITLE 38692 and 21117: dual specificity phosphatase molecules and uses
JOURNAL Patent: WO 0173059-A 1 04-OCT-2001;
therefor
Patent: Millennium Pharmaceuticals, Inc. (US)
FEATURES
source Location/Qualifiers
1..3544
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
589..2558
/note="unassigned protein product"
/codon_start=1
/protein_id="CAD10218.1"
/db_xref="GI:16509304"
/db_xref="REMBL:CAD10218"
/translation="MAHEMIGTQIVTERVALLESGETEKLIDSRPPEVYNTSHILE
AININCSKLMKRLQODKVLITELIOASHKVIDIOSGVVYDDSDOVASLSDC
FLTVLIGKLEKSPNSVHLAGAPSPRCRGICRGKSTIVPCISOPCLPVNIEPT
RILPNVYLGCGRPVNLKELMQNGICGVLANASTCRKPPPIPSHLLRPVANDPCEK
ILPWLKSVDFIEKAKANGCVLHCLAGISRATTAIAIMKMDMSIDEAYRFXE
KRPITLIPNFIQLDYEKKINOTGAGSPSKLKLHLKNEBVPVAVSEGGQSE
TPLSPCADATSEAGORPVPASVSPVPSLEDSPLVOALSGHLSDRLD
SNLKRSPSIDIKSVSASMAASLHGFSSDALFYKPSITLIDGNTKLCPSPVOE
LSRQTPETSPDKREASIPKLUOTARPSDSSKRLHVRSSSCTARSLSPHREGS
VEDNHTSFLFGJSTQOHLTKSAGJLKGWHSIDILAPORTSTSLTSSWTFATESHF
YSAIAVGGSSAYSASCSQLPTCGDQVSVRRORRPSDADRSRWHBESPEKQFK
RRSCQMEFBSIMSENRSRELKGVGSQSSFSGSEIIEVS"

ORIGIN
Query Match 94.9%; Score 1893; DB 6; Length 3544;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCCCATGAGTGAATGAGTCAAAATTTACTGAGAGGTTGGTGGCTGTGCTGGAA 60
Db 589 ATGGCCCATGAGTGAATGAGTCAAAATTTACTGAGAGGTTGGTGGCTGTGCTGGAA 648
QY 61 AGTGAACGGAAGAGTGTCTAATTGATAGCGGCAATTTGTGGAATACAAATACATCC 120
Db 649 AGTGAACGGAAGAGTGTCTAATTGATAGCGGCAATTTGTGGAATACAAATACATCC 708
QY 121 CACATTTTGAAGCATTAATATCAACTGCTCCAGGTTATGAAGCGAAGTGGCAACAG 180
Db 709 CACATTTTGAAGCATTAATATCAACTGCTCCAGGTTATGAAGCGAAGTGGCAACAG 768
QY 181 GACAAAGTGTAAATTAAGAGTCAATCCAGCATTCAGCAAAACATAAGTGTGACATTTAT 240
Db 769 GACAAAGTGTAAATTAAGAGTCAATCCAGCATTCAGCAAAACATAAGTGTGACATTTAT 828

QY 241 TGCAGTCAGAGGTTGTAGTTTACATCAAGCTCCCAAGATGTTGCTCTCTCTTCA 300
Db 829 TGCAGTCAGAGGTTGTAGTTTACATCAAGCTCCCAAGATGTTGCTCTCTCTTCA 888
QY 301 GACTGTTTTCTCAGTGTACTCTTGGGTAACTGAGAGAGCTTCAACTGTGTTCACCTG 360
Db 889 GACTGTTTTCTCAGTGTACTCTTGGGTAACTGAGAGAGCTTCAACTGTGTTCACCTG 948
QY 361 CTTGCAAGTGGGTTTGTGATGTTCTCTGTTGTTTCCCTGGCTCTGTGAGGAAATTC 420
Db 949 CTTGCAAGTGGGTTTGTGATGTTCTCTGTTGTTTCCCTGGCTCTGTGAGGAAATTC 1008
QY 421 ACTCTAGTCCCTACCTGCAATTTCTCAGCCCTTGTAACCTGTTGCCAACTGGGCCAAC 480
Db 1009 ACTCTAGTCCCTACCTGCAATTTCTCAGCCCTTGTAACCTGTTGCCAACTGGGCCAAC 1068
QY 481 CGAATTCCTCCCAATCTTTATCTTGGCTGCCAGAGATGTCTTCAACAGAGCTGATTA 540
Db 1069 CGAATTCCTCCCAATCTTTATCTTGGCTGCCAGAGATGTCTTCAACAGAGCTGATTA 1128
QY 541 CAGCAGAAATGGGATGTTATGTTAAATGCCAGTATACCTGTCCAAAGCCTGACTTT 600
Db 1129 CAGCAGAAATGGGATGTTATGTTAAATGCCAGTATACCTGTCCAAAGCCTGACTTT 1188
QY 601 ATCCCGAGTCTCATTTCTCGGCTGTGCTGCTGGAATGAGACAGCTTTGTGAGAAATTTTG 660
Db 1189 ATCCCGAGTCTCATTTCTCGGCTGTGCTGCTGGAATGAGACAGCTTTGTGAGAAATTTTG 1248
QY 661 CCGTGTGTTGACCAATCAGTATGTTTCAATTGAGAAAGCAAAAGCTTCAATGATGTTGT 720
Db 1249 CCGTGTGTTGACCAATCAGTATGTTTCAATTGAGAAAGCAAAAGCTTCAATGATGTTGT 1308
QY 721 CTAGTCACTGTTTATGCTGGGATCTTCGCGCTCGCCACCATGCTATGCTATCATG 780
Db 1309 CTAGTCACTGTTTATGCTGGGATCTTCGCGCTCGCCACCATGCTATGCTATCATG 1368
QY 781 AAGAGATGACATGCTTTAATGATGAACTTCAAGTTTGAAAGAAAGAAAGAACTTACT 840
Db 1369 AAGAGATGACATGCTTTAATGATGAACTTCAAGTTTGAAAGAAAGAAAGAAAGAACTTACT 1428
QY 841 ATATCTCCAAATCTCAATTTTCTGGGCAACTCTCTGACTATGAGAAAGATTAAAGAC 900
Db 1429 ATATCTCCAAATCTCAATTTTCTGGGCAACTCTCTGACTATGAGAAAGATTAAAGAC 1488
QY 901 CAGACTGAGCATCAGAGGCAAGAGCAAACTCAAGCTGCTGACCTGAGAAAGCAAAAT 960
Db 1489 CAGACTGAGCATCAGAGGCAAGAGCAAACTCAAGCTGCTGACCTGAGAAAGCAAAAT 1548
QY 961 GAACTGTGCTGCTGTCTCAGAGGTTGAGACAGAAAGCGCCCTCACTCAAGCCATC 1020
Db 1549 GAACTGTGCTGCTGTCTCAGAGGTTGAGACAGAAAGCGCCCTCACTCAAGCCATC 1608
QY 1021 TGTGCCGACTGTGCTACTCAGAGGAGCAGAGCAAAAGGCGGTGATCCCGGACAGT 1080
Db 1609 TGTGCCGACTGTGCTACTCAGAGGAGCAGAGCAAAAGGCGGTGATCCCGGACAGT 1668
QY 1081 CCCAGCGTCCAGCGCTGACAGCGTGTGTTAGAGACAGCCGCTGTGTAACAGCGCTTC 1140
Db 1669 CCCAGCGTCCAGCGCTGACAGCGTGTGTTAGAGACAGCCGCTGTGTAACAGCGCTTC 1728
QY 1141 AGTGGCTGCACTGTGCTGCGAGACAGGCTGGAAGACAGCAATAGCTCAAGGCTTCTTC 1200
Db 1729 AGTGGCTGCACTGTGCTGCGAGACAGGCTGGAAGACAGCAATAGCTCAAGGCTTCTTC 1788
QY 1201 TCTCTGATATCAAAATCAGTTTCAATTTGAGCAGAGATGGCAGATCTTACATGCTTC 1260
Db 1789 TCTCTGATATCAAAATCAGTTTCAATTTGAGCAGAGATGGCAGATCTTACATGCTTC 1848
QY 1261 TCTCTATCAGAAAGATGCTTTGGAATATCAAACTTTCATCTGTGATGGAGCAAC 1320
Db 1849 TCTCTATCAGAAAGATGCTTTGGAATATCAAACTTTCATCTGTGATGGAGCAAC 1908

QY 1321 AAGCTATGCGAGTTCCTCCCTGTTACAGAACTATCGAGAGAGACTCCCGAAGCAGCTT 1380
Db 1909 AAGCTATGCGAGTTCCTCCCTGTTACAGAACTATCGAGAGAGACTCCCGAAGCAGCTT 1368
QY 1381 GATPAGAGAGAGAGCAGCAGTTCCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1440
Db 1969 GATPAGAGAGAGAGCAGCAGTTCCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2028
QY 1441 AGCAGAGAGATTCGATTCGATCAGAAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
Db 2029 AGCAGAGAGATTCGATTCGATCAGAAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2088
QY 1501 TCTTCAGTGCATGAG 1560
Db 2089 TCTTCAGTGCATGAG 2148
QY 1561 CTTTTCACCGAG 1620
Db 2149 CTTTTCACCGAG 2208
QY 1621 TCGGATATCTTGGCCCCCGAG 1680
Db 2209 TCGGATATCTTGGCCCCCGAG 2268
QY 1681 ACAAGTCTCTACAGATTTCT 1740
Db 2269 ACAAGTCTCTACAGATTTCT 2328
QY 1741 GCGTACAGTGCAG 1800
Db 2329 GCGTACAGTGCAG 2388
QY 1801 CAGAGCCAG 1860
Db 2389 CAGAGCCAG 2448
QY 1861 AAGCAGATTTAAAGCAG 1920
Db 2449 AAGCAGATTTAAAGCAG 2508
QY 1921 AAGTCAAGGAG 1980
Db 2509 AAGTCAAGGAG 2568
QY 1981 ATCATTTAGAGTCTCC 1995
Db 2569 ATCATTTAGAGTCTCC 2583

RESULT 8
AF506796 3566 bp mRNA linear PRI 30-OCT-2003
LOCUS Homo sapiens dual specificity phosphatase 16 (DUSP16) mRNA,
DEFINITION complete cds.
ACCESSION AF506796
VERSION AF506796.1 GI:25573087
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 3566)
AUTHORS Hoonnaert,I., Marynen,P., Goris,J., Sciote,R. and Baens,M.
TITLE MAPK phosphatase DUSP16/MKP-7, a candidate tumor suppressor for
chromosome region 12p12-13, reduces BCR-ABL-induced transformation
JOURNAL 14586399
PUBMED 14586399
REFERENCE 2 (bases 1 to 3566)
AUTHORS Hoonnaert,I., Marynen,P. and Baens,M.
TITLE Direct Submission
JOURNAL Submitted (26-APR-2002) Department for Human Genetics-Flanders
Interuniversity Institute for Biotechnology (VIB), Katholieke
Universiteit Leuven, Herestraat 49, Leuven B-3000, Belgium

FEATURES
source location/Qualifiers
1..3566
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="12"
/map="12p13-p12"
/ligase_type="bone marrow"
1..3566
/gene="DUSP16"
/note="synonym: MKP7"
633..2630
/gene="DUSP16"
/EC_number="3.1.3.48"
/EC_number="3.1.3.16"
/codon_start=1
/product="dual specificity phosphatase 16"
/protein_id="AA075120.1"
/db_xref="GI:25573088"
/translation="MAHEMIGTOIVTERLVALLSEGREKVLIDSRPVEVNTSHILE
AINNGSKLMRRLQODKVLITELIQSHAKRVIDSQKVVVDQSDQVAVLSDD
FLTVLGLKERSFNSVHLIAGFAFSPRCFGLCEKSTLPTCI SQPLPVANIPT
RILPNLYGCDVDVINKEMQONGIGVLANSNCPDPDFIPESHFLRPVNDSECK
ILPMLDKSVDFIERAKASNGCVLVHCLAGISRATIAIAYIMKMDMSIDAYRFEK
KRPTISPENFLGOLDYERKIKNOTGASGPKSLKLIHLEKPEMPEVAVSEGGOKSE
TPLSPPCDASATSEBAAGORPVHPASVSPVQSLLEDSPLOALSGHLISADRLD
SNKLRKRSULDIKVSYSASMAASLHGFSSSEDLBYRYKRETTLDGTRKLCQSPVDE
LSBOTPETSPPKEASIPKLIQTARFSDOSKRLSHVPTSSSGTAQNSLSPLRSGS
VEDVYHNSFLPGLTSOHLTKSAGLIGKMHSDILAPOSTPBLTSSWFATRSR
YSASAIYGGASYSAYSCSLPTCGDQVYSVRROKPSRDRSRSHWHEBSPFEKOFK
RSCQMEFGESIMSENREELKVGSGSSPSGMEILEVS"

ORIGIN
Query Match 94.9%; Score 1893; DB 9; Length 3566;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGGCCCATGAGATGATGGAACCAATTGTTACTGAGAGAGTGTGGCTCTGTGGA 60
Db 633 ATGGCCCATGAGATGATGGAACCAANTTTTACTGAGAGAGTGTGGCTCTGTGGA 692
QY 61 AGTGAACGGAAGAGAGTGTGCTAATTGATGAGCCGAGCTTTGTGAATCAATACATCC 120
Db 693 AGTGAACGGAAGAGAGTGTGCTAATTGATGAGCCGAGCTTTGTGAATCAATACATCC 752
QY 121 CACATTTTGAAGCATTATATATCACTGCTCCAGCTTATGAAGCGAAGTTGCAACG 180
Db 753 CACATTTTGAAGCATTATATATCACTGCTCCAGCTTATGAAGCGAAGTTGCAACG 812
QY 181 GACAAAGTGTAAATTACAGAGCTCATCCAGATTCAGCGAAACATAAGTTGACATTGAT 240
Db 813 GACAAAGTGTAAATTACAGAGCTCATCCAGATTCAGCGAAACATAAGTTGACATTGAT 272
QY 241 TGCAGTCAGAAAGTTGATTTAGATCAAAAGCTCCCAAGATGTGCTCTCTCTTCA 300
Db 873 TGCAGTCAGAAAGTTGATTTAGATCAAAAGCTCCCAAGATGTGCTCTCTCTTCA 932
QY 301 GACTGTTTTCACATGTAATCTTGGGTAACCTGGAAGAGCTTCAACTCTGTTACCTG 360
Db 933 GACTGTTTTCACATGTAATCTTGGGTAACCTGGAAGAGCTTCAACTCTGTTACCTG 992
QY 361 CTTGCAAGTGGGTTTGTGAGATTCCTGTTGTTTCCCTGGCCCTGTGAAGAAATCC 420
Db 993 CTTGCAAGTGGGTTTGTGAGATTCCTGTTGTTTCCCTGGCCCTGTGAAGAAATCC 1052
QY 421 ACTCTAGTCCCTAACCTGCAATTTCTCAGCCTTGCTTACCTGTTGCCAACATGGGCCAAC 480
Db 1053 ACTCTAGTCCCTAACCTGCAATTTCTCAGCCTTGCTTACCTGTTGCCAACATGGGCCAAC 1112
QY 481 CGAATTTTCCCAATCTTTATCTTGGCTGCCACGAGATGCTCTCAACAGAGCTGATG 540
Db 1113 CGAATTTTCCCAATCTTTATCTTGGCTGCCACGAGATGCTCTCAACAGAGCTGATG 1172

QY 541 CAGCAGATGGAGTGGTATGTGTAAATGACAGTATACCTGTCCAAAGCCTGACTT 600
|||||
DB 1173 CAGCAGATGGAGTGGTATGTGTAAATGACAGTATACCTGTCCAAAGCCTGACTT 1232
QY 601 ATCCCGAGTCTCATTTCTGCGGTGCTGTGATGACAGCTTTTGTGAAATTTTG 660
|||||
DB 1233 ATCCCGAGTCTCATTTCTGCGGTGCTGTGATGACAGCTTTTGTGAAATTTTG 1292
QY 661 CCGTGTGTGACAAATAGTATGATTTTATGAGAAAGCAAAAGCCTTCATGATGTT 720
|||||
DB 1293 CCGTGTGTGACAAATAGTATGATTTTATGAGAAAGCAAAAGCCTTCATGATGTT 1352
QY 721 CTAGTGACGTGTTAGCTGGGATCTCCGCTCCGCGACCATCGTATGCTTACATG 780
|||||
DB 1353 CTAGTGACGTGTTAGCTGGGATCTCCGCTCCGCGACCATCGTATGCTTACATG 1412
QY 781 AAGAGATGACATGCTTTAGTGAAGCTTACAGATTTGTGAAGAAAGAAAGCCTACT 840
|||||
DB 1413 AAGAGATGACATGCTTTAGTGAAGCTTACAGATTTGTGAAGAAAGAAAGCCTACT 1472
QY 841 ATATCTCAAACTTCAATTTTCTGGGCAACTCTGACATATGAGAAAGATTAAGAAC 900
|||||
DB 1473 ATATCTCAAACTTCAATTTTCTGGGCAACTCTGACATATGAGAAAGATTAAGAAC 1532
QY 901 CAGACTGAGACATCAGGGCCAAAGAGCAACTCAAGCTGCTGACCTGAGAAAGCCAAAT 960
|||||
DB 1533 CAGACTGAGACATCAGGGCCAAAGAGCAACTCAAGCTGCTGACCTGAGAAAGCCAAAT 1592
QY 961 GAACTGTCTCTGCTCTCAGAGGGTGAAGCAAGAAAGCAGAGCCCTCAGTCCACCC 1020
|||||
DB 1593 GAACTGTCTCTGCTCTCAGAGGGTGAAGCAAGAAAGCAGAGCCCTCAGTCCACCC 1652
QY 1021 TGTGCGGACTCTGCTACTCAGAGGCGAGCAAGAAAGCCCTGCACTCCGCGAGGTT 1080
|||||
DB 1653 TGTGCGGACTCTGCTACTCAGAGGCGAGCAAGAAAGCCCTGCACTCCGCGAGGTT 1712
QY 1081 CCCAGGCTGCGCAGCGCTGAGCGGTGCTGTAGAGGACAGCCGCTGCTGACAGGGCTC 1140
|||||
DB 1713 CCCAGGCTGCGCAGCGCTGAGCGGTGCTGTAGAGGACAGCCGCTGCTGACAGGGCTC 1772
QY 1141 AGTGGGCTGACCTGTCTCCGAGCAGAGCTGGAAGACAGCAATTAAGCTCAAGCTTCTTC 1200
|||||
DB 1773 AGTGGGCTGACCTGTCTCCGAGCAGAGCTGGAAGACAGCAATTAAGCTCAAGCTTCTTC 1832
QY 1201 TCTCTGATATCAATATGATTTTATGATTTTCAAGCAGCAGTCTTCAATGCTTCC 1260
|||||
DB 1833 TCTCTGATATCAATATGATTTTATGATTTTCAAGCAGCAGTCTTCAATGCTTCC 1892
QY 1261 TCTCTATCAGAAATGCTTTGGAATCTTAAACCTTCCACTCTGATGAGGAGCCAAAC 1320
|||||
DB 1893 TCTCTATCAGAAATGCTTTGGAATCTTAAACCTTCCACTCTGATGAGGAGCCAAAC 1952
QY 1321 AAGCTATGCAAGTCTTCTCCCTGTTCAAGAACTATCGAGAGACCTCCGAAACCAAGTCT 1380
|||||
DB 1953 AAGCTATGCAAGTCTTCTCCCTGTTCAAGAACTATCGAGAGACCTCCGAAACCAAGTCT 2012
QY 1381 GATTAAGAGAGAGCAGCATCTCCCAAGAAAGCTCAGAGCCGCTTCAAGCAGCAG 1440
|||||
DB 2013 GATTAAGAGAGAGCAGCATCTCCCAAGAAAGCTCAGAGCCGCTTCAAGCAGCAG 2072
QY 1441 AGAAGAGATTTGATTCGATCGTCAAGCCAGAGAGTGGACCCGCAAGAGTCCCTTTTA 1500
|||||
DB 2073 AGAAGAGATTTGATTCGATCGTCAAGCCAGAGAGTGGACCCGCAAGAGTCCCTTTTA 2132
QY 1501 TCTCACTGATGAGATGAGGAGGCTGAGAGCAATTAACACACAGCTTCTTTTCGCG 1560
|||||
DB 2133 TCTCACTGATGAGATGAGGAGGCTGAGAGCAATTAACACACAGCTTCTTTTCGCG 2192
QY 1561 CTTTCCACAGCCAGCAGCACTTCAAGAGTCTGCTGCGCTGAGGCTTAAAGGCTGGCAC 1620
|||||
DB 2193 CTTTCCACAGCCAGCAGCACTTCAAGAGTCTGCTGCGCTTAAAGGCTGGCAC 2252
QY 1621 TCGGATATCTTGGCCCCCAGACCTTACCCCTTCCCTGACAGCAGCTGATTTTGGC 1680

DB 2253 TCGGATATCTTGGCCCCCAGACCTTACCCCTTCCCTGACAGCAGCTGATTTTGGC 2312
QY 1681 ACAGAGTCTTACACTTCTTACTGCTTCAAGCATTTAGAGAGCAGTCCAGTACTCT 1740
DB 2313 ACAGAGTCTTACACTTCTTACTGCTTCAAGCATTTAGAGAGCAGTCCAGTACTCT 2372
QY 1741 GCTTACAGTCCAGCAGCAGTCCCACTTCCGAGACCAAGTCTTCTGAGCAGGCGG 1800
DB 2373 GCTTACAGTCCAGCAGCAGTCCCACTTCCGAGACCAAGTCTTCTGAGCAGGCGG 2432
QY 1801 CAGAGCCAAAGTGAAGCAGACTGCTGCGCGGAGCTGAGCAATGAAGAGCCCTTTGAA 1860
DB 2433 CAGAGCCAAAGTGAAGCAGACTGCTGCGCGGAGCTGAGCAATGAAGAGCCCTTTGAA 2492
QY 1861 AAGCAGTTTAAACGAGAAAGCTGCCAAATGGAATTTGAGAGAGCATATGTCAGAGAAC 1920
DB 2493 AAGCAGTTTAAACGAGAAAGCTGCCAAATGGAATTTGAGAGAGCATATGTCAGAGAAC 2552
QY 1921 AGGTCAAGGGAAGAGCTGGGGAAGAGGAGCAGTCAAGCTTCTTCCGCGCAGCATGGA 1980
DB 2553 AGGTCAAGGGAAGAGCTGGGGAAGAGGAGCAGTCAAGCTTCTTCCGCGCAGCATGGA 2612
QY 1981 ATCATTTAGAGTCTCC 1995
DB 2613 ATCATTTAGAGTCTCC 2627

RESULT 9
AX374994 3766 bp DNA linear PAT 01-MAR-2002
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1
Tang,Y.T., Elliott,V.S., Rankkumar,J., Yao,M.G., Buford,N.,
Wang,Y.E., Stewart,E.A., Gandhi,A.R., Peterson,C., Lee,E.A.,
Hafalla,A.J., Lu,D.A., Tribouley,C.M., Griffin,J.A., Baughn,M.R.,
Yue,H., Warren,B.A., Nguyen,D.B. and Walla,N.K.
Protein phosphatases
Patent: WO 0210363-A 17 07-FEB-2002;
JOURNAL
Incyte Genomics, Inc. (US)
FEATURES
Location/Qualifiers
source
1..3766
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="Incyte ID No: 7480570CB1"

ORIGIN
Query Match 94.9%; Score 1893; DB 6; Length 3766;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGGCCATGAGATGATTTGAACTCAATTTTACTGAGAGGTGGCTGCTGGAA 60
DB 538 ATGGCCATGAGATGATTTGAACTCAATTTTACTGAGAGGTGGCTGCTGGAA 597
QY 61 AGTGAACGGAAGAAAGTGTCTAATTTAGTACGGGCAATTTGTGAAATACATATCC 120
DB 598 AGTGAACGGAAGAAAGTGTCTAATTTAGTACGGGCAATTTGTGAAATACATATCC 657
QY 121 CACATTTTGAAGCCATTAATATCACTGCTCAAGCTTATGAGGAAAGTTGCAACG 180
DB 658 CACATTTTGAAGCCATTAATATCACTGCTCAAGCTTATGAGGAAAGTTGCAACG 717
QY 181 GACAAAGTGTATTACAGAGCTCATCCAGCATTTGAGGAAACATTAAGTTGACATTTGAT 240

Db 718 GACAAAGTGAATTAACAGAGCTCATCCAGCATTCAGGAAACATAAGGTTGACATTGAT 777
Qy 241 TGCAGTGAAGAGTGAATTAAGATTAAGTAAGCTCCCAAGTGTGCTCTCTCTTCA 300
Db 778 TGCAGTGAAGAGGTTGATTTAGATCAAGCTCCCAAGATGTGCTCTCTCTCTTCA 837
Qy 301 GACTGTTTCTCAGTGAATCTGGGTAAGCTGGAAGAGCTTCACTGTTCACTG 360
Db 838 GACTGTTTCTCAGTGAATCTGGGTAAGCTGGAAGAGCTTCACTGTTCACTG 897
Qy 361 CTTCAGAGTGGGTTTGTGAGTCTCTGCTGTTTCCCTGGCTCTGTGAAGAAAAATCC 420
Db 898 CTTCAGAGTGGGTTTGTGAGTCTCTGCTGTTTCCCTGGCTCTGTGAAGAAAAATCC 957
Qy 421 ACTGATGCTCCATCTGATTTCTCAGCTTGTCTTACTGTTGCAACATTTGGGCCAAC 480
Db 958 ACTGATGCTCCATCTGATTTCTCAGCTTGTCTTACTGTTGCAACATTTGGGCCAAC 1017
Qy 481 CGAATTTCTCCAACTTTTATCTTGGCTGCCAGAGATGTCTCAACAGAGCTGATA 540
Db 1018 CGAATTTCTCCAACTTTTATCTTGGCTGCCAGAGATGTCTCAACAGAGCTGATA 1077
Qy 541 CAGAGATAGGAGTGGTATGTGTTAATGCAAGCTATACTGTCCAAAAGCTGACT 600
Db 1078 CAGAGATAGGAGTGGTATGTGTTAATGCAAGCTATACTGTCCAAAAGCTGACT 1137
Qy 601 ATCCCGAGTCTGATTTCCGCGTGGCTGTGAATGACAGCTTTTGTGAAGAAATTTTG 660
Db 1138 ATCCCGAGTCTGATTTCCGCGTGGCTGTGAATGACAGCTTTTGTGAAGAAATTTTG 1197
Qy 661 CCGTGGTTGGAACAATCAGTAGATTTCAATTGAGAAAACAAGCTCAATGATGTGT 720
Db 1198 CCGTGGTTGGAACAATCAGTAGATTTCAATTGAGAAAACAAGCTCAATGATGTGT 1257
Qy 721 CTGATGACCTGTTAGTGGGATCTCCGCTCCGCCACATGCTATTCGCTACATCATG 780
Db 1258 CTGATGACCTGTTAGTGGGATCTCCGCTCCGCCACATGCTATTCGCTACATCATG 1317
Qy 781 AAGAGATGACATGCTTGTAGATGAAGCTTACAGATTTGTGAAGAAAAAGACCTACT 840
Db 1318 AAGAGATGACATGCTTGTAGATGAAGCTTACAGATTTGTGAAGAAAAAGACCTACT 1377
Qy 841 ATATCTCCAACTTCAATTTTCTGGGCCAACTCTGAGCTATGAGAAAGATTAGAAG 900
Db 1378 ATATCTCCAACTTCAATTTTCTGGGCCAACTCTGAGCTATGAGAAAGATTAGAAG 1437
Qy 901 CAGACTGAGACATCAGAGGCCAAAGACAACTAACCTGTGCACTTGAAGAACCAAT 960
Db 1438 CAGACTGAGACATCAGAGGCCAAAGACAACTAACCTGTGCACTTGAAGAACCAAT 1497
Qy 961 GAACCTGTCCCTGCTGTCTCAGAGGGTGGACAGAAAAGGACGCCCTCAGTCCACCC 1020
Db 1498 GAACCTGTCCCTGCTGTCTCAGAGGGTGGACAGAAAAGGACGCCCTCAGTCCACCC 1557
Qy 1021 TGTGCCGACTGTGCTAAGCTTCAAGAGGACAGAAAGGCCCTGTGATCCCGCAGCGTG 1080
Db 1558 TGTGCCGACTGTGCTAAGCTTCAAGAGGACAGAAAGGCCCTGTGATCCCGCAGCGTG 1617
Qy 1081 CCAGAGTGGCCAGAGCGGTGAGCGCTGCTTGAAGAGAGCGCCGCTGTATAGCGCGTCC 1140
Db 1618 CCAGAGTGGCCAGAGCGGTGAGCGCTGCTTGAAGAGAGCGCCGCTGTATAGCGCGTCC 1677
Qy 1141 AGTGGGCTGACCTGTCCGACAGACAGGCTGGAAGACAGCATTAAGCTTAAGGCTTCCTTC 1200
Db 1678 AGTGGGCTGACCTGTCCGACAGACAGGCTGGAAGACAGCATTAAGCTTAAGGCTTCCTTC 1737
Qy 1201 TCTCTGATATCAAAATCAATTCATATTCAGCCAGCATGCGCATCTTACATGCTTC 1260
Db 1738 TCTCTGATATCAAAATCAATTCATATTCAGCCAGCATGCGCATCTTACATGCTTC 1797
Qy 1261 TCCCTATCAGAGATGCTTTGGAATATCAAAACCTTCCATCTCTGTGATGGACCAAC 1320
Db 1798 TCCCTATCAGAGATGCTTTGGAATATCAAAACCTTCCATCTCTGTGATGGACCAAC 1857

Qy 1321 AAGCTATGCGCAAGTCTCCCTGTTCAGGAATATCGAGACAGCTCCCGAAACAAGTCT 1380
Db 1858 AAGCTATGCGCAAGTCTCCCTGTTCAGGAATATCGAGACAGCTCCCGAAACAAGTCT 1917
Qy 1381 GATTAAGGAGGAGGACGACATCCCGAAGAAAGCTGACACCGGCTTCAAGAGCCGAG 1440
Db 1918 GATTAAGGAGGAGGACGACATCCCGAAGAAAGCTGACACCGGCTTCAAGAGCCGAG 1977
Qy 1441 AGCAAGCATTCATTCGCTGAGAACCAAGACAGTGGACACCGCCAGAGGCTCCCTTTTA 1500
Db 1978 AGCAAGCATTCATTCGCTGAGAACCAAGACAGTGGACACCGCCAGAGGCTCCCTTTTA 2037
Qy 1501 TCTCCATGCTCATCGAAGTGGAGCGCTGAGGACAAATTACACACAGCTTCTTTTGGCC 1560
Db 2038 TCTCCATGCTCATCGAAGTGGAGCGCTGAGGACAAATTACACACAGCTTCTTTTGGCC 2097
Qy 1561 CTTTCACACGACGACAGACCTCAAGAGTGTGTGCGCTTGAAGGCTTGAAGGCTGGCAC 1620
Db 2098 CTTTCACACGACGACAGACCTCAAGAGTGTGTGCGCTTGAAGGCTTGAAGGCTGGCAC 2157
Qy 1621 TCGAATATCTTGGCCCCCAGACCTCTACCCCTTCCCTGACAGACGCTGATTTTGGC 1680
Db 2158 TCGAATATCTTGGCCCCCAGACCTCTACCCCTTCCCTGACAGACGCTGATTTTGGC 2217
Qy 1681 ACAGAGTCTCTACATTTCTACTGTGCTGACGCATCTACGAGAGGACGTGCCAGTTACT 1740
Db 2218 ACAGAGTCTCTACATTTCTACTGTGCTGACGCATCTACGAGAGGACGTGCCAGTTACT 2277
Qy 1741 GCTTACAGCTGAGCAGCTGCCCTGAGGAGACAAAGCTATTCGTGTCGAGAGGCGG 1800
Db 2278 GCTTACAGCTGAGCAGCTGCCCTGAGGAGACAAAGCTATTCGTGTCGAGAGGCGG 2337
Qy 1801 CAGAAAGCAAGTGAAGAGCTGACTCGCGGCGAGCTGGACATGAAGAGAGCCCTTTGAA 1860
Db 2338 CAGAAAGCAAGTGAAGAGCTGACTCGCGGCGAGCTGGACATGAAGAGAGCCCTTTGAA 2397
Qy 1861 AAGCATTTAAACGAGAAAGCTGCGCAATATGAAATTTGAGAGAGCATATCTCAGAGAAC 1920
Db 2398 AAGCATTTAAACGAGAAAGCTGCGCAATATGAAATTTGAGAGAGCATATCTCAGAGAAC 2457
Qy 1921 AGGTACCGGAGAGAGCTGGGAAAGTGGGAGTCAAGTCTAGCTTTTGGGAGCAGATGAA 1980
Db 2458 AGGTACCGGAGAGAGCTGGGAAAGTGGGAGTCAAGTCTAGCTTTTGGGAGCAGATGAA 2517
Qy 1981 ATCATTTAGGTCTCC 1995
Db 2518 ATCATTTAGGTCTCC 2532

RESULT 10
BD171157
LOCUS BD171157 4790 bp DNA linear PAT 17-JAN-2003
DEFINITION Novel gene and protein encoded thereby.
ACCESSION BD171157.1 GI:27876969
VERSION BD171157.1
KEYWORDS WO 02052005-A/13.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 4790)
Ohara,O., Nagase,T. and Nakajima,D.
Novel gene and protein encoded thereby
Patent: WO 02052005-A 13 04-JUL-2002;
KAZUSA DNA RESEARCH INSTITUTE,OSAMU OHARA,TAKAHIRO NAGASE, DAISUKE
NAKAJIMA
OS Homo sapiens (human)
PN WO 02052005-A/13
PD 04-JUL-2002
PF 20-DEC-2001 WO 2001JP011217
PI 22-DEC-2000 JP 00P 389742
PI OSAMU OHARA,TAKAHIRO NAGASE,DAISUKE NAKAJIMA

	CL2M15/12, C07K14/47	Novel gene and protein encoded thereby	Location/Qualifiers
PC			
CC	Novel gene and protein encoded thereby		
FH	Key		
FT	CDS		(184). (2178).

FEATURES	location/Qualifiers
source	1..4790
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"

ORIGIN

Query Match	94.9%	Score 1893;	DB 6;	Length 4790;
Best Local Similarity	99.9%	Pred. No. 0;		
Matches 1993; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0

Oy	1	TTGGCCCATGAGATGATATTGGAACTCAAAATTTGTTACTGAGAGGGTGGGTGGCTCTGCTGGAA	60
Db	184	ATGGCCCATGAGATGATATTGGAACTCAAAATTTGTTACTGAGAGGGTGGGTGGCTCTGCTGGAA	243
Oy	61	AGTGAAACGAAAAAAGTCGTCTAATTGATAGACGGACCATTTGTGGAAATACAATTCATCC	120
Db	244	ACTGGAACGGAAAAAAGTCTGCTAATTGATAGACGGACCATTTGTGGAAATACAATTCATCC	303
Oy	121	CACATTTTGGAAAGCCATTAAATATCAACTGCTCCAAAGCTTATGAAACGAAAGTTTGAACAG	180
Db	304	CACATTTTGGAAAGCCATTAAATATCAACTGCTCCAAAGCTTATGAAACGAAAGTTTGAACAG	363
Oy	181	GACAAAGTCTTAATTACAGAGCTCATCCAGCAATTAGCGAAACATAAGTTAGCAATTGAT	240
Db	364	GACAAAGTCTTAATTACAGAGCTCATCCAGCAATTAGCGAAACATAAGTTAGCAATTGAT	423
Oy	241	TGCAGTCAGAAGGTGTGATTTACATCAAAAGCTCCCAAGATGTTGGCCTCTCTCTTCA	300
Db	424	TGCAGTCAGAAGGTGTGATTTACATCAAAAGCTCCCAAGATGTTGGCCTCTCTCTTCA	483
Oy	301	GACTGTTTTCTCACTGTACTTCGTGGGTAACTGGAAAGGCTTCAACTGTTCACCTG	360
Db	484	GACTGTTTTCTCACTGTACTTCGTGGGTAACTGGAAAGGCTTCAACTGTTCACCTG	543
Oy	361	CTTGGAGAGTGGTGTGGCTGAGTTCCTCTGTTTCCCTGGCCTCTGTGGAAGAAATCC	420
Db	544	CTTGGAGAGTGGTGTGGCTGAGTTCCTCTGTTTCCCTGGCCTCTGTGGAAGAAATCC	603
Oy	421	ACTAGTACCCCACTGACATTTCTCAGCCTTGCTTAACCTGTGGCCAACTTTGGGCCAAC	480
Db	604	ACTAGTACCCCACTGACATTTCTCAGCCTTGCTTAACCTGTGGCCAACTTTGGGCCAAC	663
Oy	481	CGAATTCCTCCCAATCTTATCTTGGCTGCGACGAGATGTCCTCAACAAAGAGCTGATA	540
Db	664	CGAATTCCTCCCAATCTTATCTTGGCTGCGACGAGATGTCCTCAACAAAGAGCTGATG	723
Oy	541	CAGCAGAAATGGATTTGGTTATGTGTTAAATGCGACGATATCTGTGTCCAAAGCTGACTTT	600
Db	724	CAGCAGAAATGGATTTGGTTATGTGTTAAATGCGACGATATCTGTGTCCAAAGCTGACTTT	783
Oy	601	ATCCCGGAGTCTCATTTCTCGCGTGGCCGTGGAATGACAGCTTTTGGAGAAATTTTGG	660
Db	784	ATCCCGGAGTCTCATTTCTCGCGTGGCCGTGGAATGACAGCTTTTGGAGAAATTTTGG	843
Oy	661	CCGTGGTTGACAAATCAGTAGATTTCATTTGAGAAAGCAAAAGCCTCCAAATGATGTGTT	720
Db	844	CCGTGGTTGACAAATCAGTAGATTTCATTTGAGAAAGCAAAAGCCTCCAAATGATGTGTT	903
Oy	721	CTAGTGCACACTGTTTAGCTGGGATCTCCCGCTCCGCGACCATGCGCTATACATCATG	780
Db	904	CTAGTGCACACTGTTTAGCTGGGATCTCCCGCTCCGCGACCATGCGCTATACATCATG	963
Oy	781	AAGAGGATGGAATGCTCTTTAGATGAAAGCTTACAGATTTGTGAAAGAAAAAGAACTACT	840
Db	964	AAGAGGATGGAATGCTCTTTAGATGAAAGCTTACAGATTTGTGAAAGAAAAAGAACTACT	1022
Oy	841	ATATCTCCAAACTTCATTTTCTGGGCCCACTCTGAGCTATGAGAAAGAAATTAAGAC	900

Db	1024	ATATCTCCAAACTTCATATTTTCTGGGGCAACTCTGTGGATCTATGAGAAAGATTAAAGAC	1083
Qy	901	CAGACTGAGCATCAGGAGCCAAAGAGCAAACTCAAGCTGTGCACCTGGAGAACCAAT	960
Db	1084	CAGACTGGAGCATCAGGAGCCAAAGAGCAAACTCAAGCTGTGCACCTGGAGAACCAAT	1143
Qy	961	GAACCTGTCCCTGTGTCTCAGAGGGTGGACAGAAAACGAGACGCCCTCAAGTCCACC	1020
Db	1144	GAACCTGTCCCTGTGTCTCAGAGGGTGGACAGAAAACGAGAGGCCCTCAAGTCCACC	1203
Qy	1021	TGTGCCCACTGTGTACTCTCAGAGGCGAGAGCAAAAGGCCGTGCATCCCGCAGAGGTG	1080
Db	1204	TGTGCCCACTGTGTACTCTCAGAGGCGAGAGCAAAAGGCCGTGCATCCCGCAGAGGTG	1263
Qy	1081	CCGAGCGTGCCGACGCTGCAGCCGCTGTCTTAAAGGACAGCCCGCTGTATCAGAGCGCTC	1140
Db	1264	CCGAGCGTGCCGACGCTGCAGCCGCTGTCTTAAAGGACAGCCCGCTGTATCAGAGCGCTC	1323
Qy	1141	AGTGGGCTGCACCTGTCCGAGACAGGCTGGAAAGACAGCAATAGCTCAAGCGTTCTTC	1200
Db	1324	AGTGGGCTGCACCTGTCCGAGACAGGCTGGAAAGACAGCAATAGCTCAAGCGTTCTTC	1383
Qy	1201	TCTCGGATCAATCAGTTTCATTTTCAAGCCGACATGGCAGATCCTTACATGGCTTC	1260
Db	1384	TCTCGGATCAATCAGTTTCATTTTCAAGCCGACATGGCAGATCCTTACATGGCTTC	1443
Qy	1261	TCCTCATCAGAAAGTGTGGTAATCTAGCAAACTTCCACTACTCTGTGATGGGACCAAC	1320
Db	1444	TCCTCATCAGAAAGTGTGGTAATCTAGCAAACTTCCACTACTCTGTGATGGGACCAAC	1503
Qy	1321	AAGCTATGCCAGTTCTCCCTCTGTTCAGAGAACTATCGAGACAGACTCCGAAACCAAGTCT	1380
Db	1504	AAGCTATGCCAGTTCTCCCTCTGTTCAGAGAACTATCGAGACAGACTCCGAAACCAAGTCT	1563
Qy	1381	GATTAAGAGGAGGACATCCCCAAAGCTGCAGACCGCGGCTTGAACAGGACAG	1440
Db	1564	GATTAAGAGGAGGACATCCCCAAAGCTGCAGACCGCGGCTTGAACAGGACAG	1623
Qy	1441	AGCAAGCGATTTGCACTTGGCTCAGAACCAAGACAGACAGCCGCGCAGAGGTCCCTTTTA	1500
Db	1624	AGCAAGCGATTTGCACTTGGCTCAGAACCAAGACAGACAGCCGCGCAGAGGTCCCTTTTA	1683
Qy	1501	TCTCOACTGCATCGAAGTGGGAGCGGTGAGAGCAATTACACACACAGCTTCTTTTGGC	1560
Db	1684	TCTCOACTGCATCGAAGTGGGAGCGGTGAGAGCAATTACACACACAGCTTCTTTTGGC	1743
Qy	1561	CTTTCCACACGCCAGACACACTTACAGAAAGTCTGTCTGGCTTGAAGGCTGGCAC	1620
Db	1744	CTTTCCACACGCCAGACACACTTACAGAAAGTCTGTCTGGCTTGAAGGCTGGCAC	1803
Qy	1621	TCGGATATCTTGGAGCCCCCAGACCTCTACCCCTTCCCTGACACAGCACTGGTATTTTGGC	1680
Db	1804	TCGGATATCTTGGAGCCCCCAGACCTCTACCCCTTCCCTGACACAGCACTGGTATTTTGGC	1863
Qy	1681	ACAGAGTCTTCACACTTCTACTCTGTCCCTCAAGCCAATTACGAGGCGAGTGCACGTTACTCT	1740
Db	1864	ACAGAGTCTTCACACTTCTACTCTGTCCCTCAAGCCAATTACGAGGCGAGTGCACGTTACTCT	1923
Qy	1741	GCTTACAGCTGCAGCCAGCTGCGCACTTGGGGAGAACCAAGTCTATTTCTGTGGCGAGCGG	1800
Db	1924	GCTTACAGCTGCAGCCAGCTGCGCACTTGGGGAGAACCAAGTCTATTTCTGTGGCGAGCGG	1983
Qy	1801	CAGAAAGCCAAGTGCAGAGCTGCACTGCGGGCGAGCTGGCATGAAGAGAGACCCCTTTGAA	1860
Db	1984	CAGAAAGCCAAGTGCAGAGCTGCACTGCGGGCGAGCTGGCATGAAGAGAGACCCCTTTGAA	2043
Qy	1861	AAGCAATTAAACGCAGAAAGCTGCCAAATTTTGGAGAGAGCATCATGTCCAGAGAAC	1920
Db	2044	AAGCAATTAAACGCAGAAAGCTGCCAAATTTTGGAGAGAGCATCATGTCCAGAGAAC	2103
Qy	1921	AGGTTCACGGGAAAGCTGGGGAAAGTGGGCACTCAAGTCTAGCTTTTGGGCGAGCATGGAA	1980
Db	2104	AGGTTCACGGGAAAGCTGGGGAAAGTGGGCACTCAAGTCTAGCTTTTGGGCGAGCATGGAA	2163

Db 1684 TCTCCACTGCATCGAAGTGGAGCGTGGAGAGCAATTACACACAGCTTCTTTTCGGC 1743

QY 1561 CTTTCCACCCGACGACGACCTTCAGGAATCTGCTGGCTTGAAGGCTTGGCAC 1620

Db 1744 CTTTCCACCCGACGACGACCTTCAGGAATCTGCTGGCTTGAAGGCTTGGCAC 1803

QY 1621 TCGGATATCTTTGGCCCCCAGACCTCTACCCCTTCCCTGACGACAGCTGTATTTTGGC 1680

Db 1804 TCGGATATCTTTGGCCCCCAGACCTCTACCCCTTCCCTGACGACAGCTGTATTTTGGC 1863

QY 1681 ACAGAGTCTCTACACTTCTACTCTGCTTCAGCATCTACGAGGACAGTGGCACTTACTCT 1740

Db 1864 ACAGAGTCTCTACACTTCTACTCTGCTTCAGCATCTACGAGGACAGTGGCACTTACTCT 1923

QY 1741 GCTTACAGCTGACGACGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1800

Db 1924 GCTTACAGCTGACGACGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1983

QY 1801 CAGAGGCAAGTACGACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1860

Db 1984 CAGAGGCAAGTACGACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2043

QY 1861 AACGAGTTTAAAGCAGAGAGCTCCAAATGGAATTTGAGAGAGCATCATGTGAGAGAC 1920

Db 2044 AACGAGTTTAAAGCAGAGAGCTCCAAATGGAATTTGAGAGAGCATCATGTGAGAGAC 2103

QY 1921 AGGTCAAGGAGAGAGCTGGGGAAGTGGGAGTCACTGCTTTTGGGACAGATGGA 1980

Db 2104 AGGTCAAGGAGAGAGCTGGGGAAGTGGGAGTCACTGCTTTTGGGACAGATGGA 2163

QY 1981 ATGATTGAGGCTTCC 1995

Db 2164 ATGATTGAGGCTTCC 2178

RESULT 12

LOCUS AB051487 4790 bp mRNA linear PRI 07-FEB-2001

DEFINITION Homo sapiens mRNA for KIAA1700 protein, partial cds.

ACCESSION AB051487

VERSION AB051487.1 GI:12697944

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases) Nagase, T., Kikuno, R., Hattori, A., Kondo, Y., Okumura, K. and Ohara, O. Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

TITLE Prediction of the coding sequences of unidentified human genes. XIX. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro

JOURNAL DNA Res. 7 (6), 347-355 (2000)

MEDLINE 21082932

PUBMED 11214970

REFERENCE 2 (bases 1 to 4790)

AUTHORS Ohara, O., Nagase, T. and Kikuno, R.

TITLE Direct Submission

JOURNAL Submitted (22-NOV-2000) Osamu Ohara, Kazusa DNA Research Institute, Department of Human Gene Research, 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:cdnaif@kazusa.or.jp, URL: http://www.kazusa.or.jp/huge, Tel: 81-438-52-3913, Fax: 81-438-52-3914)

FEATURES

source

1. .4790

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="FJ15353"

/note="vector: pBluescript SK plus"

1. .4790

/gene="KIAA1700"

<109. .2181

/gene="KIAA1700"

/note="Start codon is not identified."

/codon_start=1

/product="KIAA1700 protein"

/protein_id="BAB21791.1"

/db_xref="GI:12697945"

/translation="AHEHTTITSWQIKGGGKRGILVMAHEMIGTOIVTERVALL
ESGTEKVLIDSRPVEVNTSHILAININCKLMKRLQDQKVLITELIOHAKKV
DIDGOKVVVPOSSODVASLSDCLITLGLKESFNSVLHAGFPAPFSPFGL
CEGKSTVPTCSISORCLPVANIGPRLIPNLVIGCORVLANKEIMQONIGVLAN
TCKPDPFIPESHFLRVNDSPCEKILPMLDSDVDEIRAKSNQCVLVHCLAGIRS
ATIALAYIMKRMDSIDEAYRVFKEKRPISNPNFLGQLDEKIKNOTAGSGKS
KIKLHLKEPNEPVPASVEGQKSETPSPCADSPATSEAGQRPVHPASVSPVQ
PILBESPLVQALSGHLSDRLSDSNKLSRFSFLDIKSVYSASVAAALHGSSED
ALEYKPSSTTLTGTHKLCOPSPVOLESDPTSPSPKEKASI PKLQTPRSPSOKR
LHSVTSSEGTQRLSLSPHRSQVEDVTHYSPLESTSOQHLTKSGCLKQHS
DLAPOTSTPSLTSWTFATESHFTSBAITLGGASVATYSCSOLPTGDOVYSVR
RQKPSDRSRSRSMHESPFEKQFRKRSQMEFESIMSENRREBLRGVSQSFSFG
SMEILEVS"

ORIGIN

Query Match 94.9%; Score 1893; DB 9; Length 4790;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCCCATGAGATGATGGAATCTCAATTTGTTACTGAGAGGTGGCTGCTGAGAA 60

Db 184 ATGGCCCATGAGATGATGGAATCTCAATTTGTTACTGAGAGGTGGCTGCTGAGAA 243

QY 61 AGTGAACGGAAAAAGTGTCTGCTAATTGATAGCCGGCCATTGTGGAATACATATATCC 120

Db 244 AGTGAACGGAAAAAGTGTCTGCTAATTGATAGCCGGCCATTGTGGAATACATATATCC 303

QY 121 CACATTTTGAAGCCATTATATCACTGCTCCAGCTTATGAGCGAAGGTGGCAACG 180

Db 304 CACATTTTGAAGCCATTATATCACTGCTCCAGCTTATGAGCGAAGGTGGCAACG 363

QY 181 GACAAAGTATTAATACAGAGCTCAATCCAGCATTCAGCGAAACATTAAGGTGATGAT 240

Db 364 GACAAAGTATTAATACAGAGCTCAATCCAGCATTCAGCGAAACATTAAGGTGATGAT 423

QY 241 TGCAGTCAGAAAGTGTAGTTTACATCAATCAAGCTCCCAAGATGTTCCCTCTCTTCA 300

Db 424 TGCAGTCAGAAAGTGTAGTTTACATCAATCAAGCTCCCAAGATGTTCCCTCTCTTCA 483

QY 301 GACTGTTTCTCACTGTAATCTTGGGTAATCTGAGAGAGAGCTTCAATCTGTTCACTG 360

Db 484 GACTGTTTCTCACTGTAATCTTGGGTAATCTGAGAGAGAGCTTCAATCTGTTCACTG 543

QY 361 CTTGCAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 420

Db 544 CTTGCAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 603

QY 421 ACTCTAGTCCCTACCTGCAATTTCTCAGCTTGTCTGCTTCACTGTTGCCAATTTGGGCCA 480

Db 604 ACTCTAGTCCCTACCTGCAATTTCTCAGCTTGTCTGCTTCACTGTTGCCAATTTGGGCCA 663

QY 481 CGAATTTCTCCCAATCTTATCTTGGCTGCCAGCGAGATGCTCTCAACAGAGCTGATA 540

Db 664 CGAATTTCTCCCAATCTTATCTTGGCTGCCAGCGAGATGCTCTCAACAGAGCTGATA 723

QY 541 CAGCAAAATGGGAATGGTATGTTAAATGACAGTATACCTGTCCTCAAGCTGACTT 600

Db 724 CAGCAAAATGGGAATGGTATGTTAAATGACAGTATACCTGTCCTCAAGCTGACTT 783

QY 601 ATCCCGAGTCTCATTTCTGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660

Db 784 ATCCCGAGTCTCATTTCTGCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 843

QY 661 CCGTGTGTGACCAATCAGTATGTTTCAATGAGAAAGCAAAAGCTTCAATGATGATGTT 720

Db 844 CCGTGTGTGACCAATCAGTATGTTTCAATGAGAAAGCAAAAGCTTCAATGATGATGTT 903

QY 721 CTAGTGCACCTGTTTACTGGGATCTCCGCTCCGCCACATGCTATCGCTTACATCATG 780

Db 904 CTAGTGCATCTGTTAGTGGATCTCCGCTCCGCACTATGCTTATCATCATG 963
Qy 781 AAGAGAGTGAACATGCTTTAGTATGACCTTACAGATTTTGGAAAAAGAACTTACT 840
Db 964 AAGAGAGTGAACATGCTTTAGTATGACCTTACAGATTTTGGAAAAAGAACTTACT 1023
Qy 841 ATATCTCCAACTTCAATTTTCTGGGCCAATCTCTGACATATGAGAAAGATTAAAGAC 900
Db 1024 ATATCTCCAACTTCAATTTTCTGGGCCAATCTCTGACATATGAGAAAGATTAAAGAC 1083
Qy 901 CAGACTGAGACATCAGGGCCAAAGACAACTCAAGCTGTGCACTGAGAAAGCCAAAT 960
Db 1084 CAGACTGAGACATCAGGGCCAAAGACAACTCAAGCTGTGCACTGAGAAAGCCAAAT 1143
Qy 961 GAACCTGCTCCTGCTGCTCAGAGGGTGGACAGAAAAAGCAGAGAGCCCTCAGTCCACC 1020
Db 1144 GAACCTGCTCCTGCTGCTCAGAGGGTGGACAGAAAAAGCAGAGAGCCCTCAGTCCACC 1203
Qy 1021 TGTGCCGACTCTGCTTACCTCAGAGGACAGACAAAGGCCCGTGCATCCCGCAGGTG 1080
Db 1204 TGTGCCGACTCTGCTTACCTCAGAGGACAGACAAAGGCCCGTGCATCCCGCAGGTG 1263
Qy 1081 CCGAGCGTCCGACGCTGCTGCTCAGAGGGTGGACAGAAAAAGCAGAGAGCCCTCAGTCCACC 1140
Db 1264 CCGAGCGTCCGACGCTGCTGCTCAGAGGGTGGACAGAAAAAGCAGAGAGCCCTCAGTCCACC 1323
Qy 1141 AGTGGGCTGACCTGCTCAGAGGACAGAGCTGAGAGACAAATTAAGCTCAAGGCTTCTTC 1200
Db 1324 AGTGGGCTGACCTGCTCAGAGGACAGAGCTGAGAGACAAATTAAGCTCAAGGCTTCTTC 1383
Qy 1201 TCTCTGATATCAATATGATTCTATTCAGCCAGCATGAGCATCTTACATGCTTC 1260
Db 1384 TCTCTGATATCAATATGATTCTATTCAGCCAGCATGAGCATCTTACATGCTTC 1443
Qy 1261 TCTCTCATCAAGATGCTTGGAACTTACAACTTCCATCTCTGATGAGACCAAC 1320
Db 1444 TCTCTCATCAAGATGCTTGGAACTTACAACTTCCATCTCTGATGAGACCAAC 1503
Qy 1321 AAGCTATGCCAGTTCTCCCTGTTTCAGAACTATCCGAGCAGACTCCGAAACCAAGTCT 1380
Db 1504 AAGCTATGCCAGTTCTCCCTGTTTCAGAACTATCCGAGCAGACTCCGAAACCAAGTCT 1563
Qy 1381 GATTAAGAGAGAACCCAGCATCTCCCAAGAAAGCTGACAGCCGACCTTTCAGACAGCCAG 1440
Db 1564 GATTAAGAGAGAACCCAGCATCTCCCAAGAAAGCTGACAGCCGACCTTTCAGACAGCCAG 1623
Qy 1441 AGCAAGCATTTGATTCGGTTCAGAAACAGAGAGAGGAGACCGCCAGAGGTCCTTTT 1500
Db 1624 AGCAAGCATTTGATTCGGTTCAGAAACAGAGAGAGGAGACCGCCAGAGGTCCTTTT 1683
Qy 1501 TCTCCACTGATTCGAAGTGGAGCGTGAAGACAATTACACACAGCTTCTTTTCCGAC 1560
Db 1684 TCTCCACTGATTCGAAGTGGAGCGTGAAGACAATTACACACAGCTTCTTTTCCGAC 1743
Qy 1561 CTTTCCACAGAGCAGAGACCTTCAAGAACTCTGCTGCTGCTTGAAGGCTTGGAC 1620
Db 1744 CTTTCCACAGAGCAGAGACCTTCAAGAACTCTGCTGCTGCTTGAAGGCTTGGAC 1803
Qy 1621 TCGGATATCTGAGCCGCCAGACCTTCAACCTTCCCTGACAGAGAGCTGATTTTCC 1680
Db 1804 TCGGATATCTGAGCCGCCAGACCTTCAACCTTCCCTGACAGAGAGCTGATTTTCC 1863
Qy 1681 AAGAGTCTTCACTTCTTCTGCTGACCATCTTCAAGAGAGAGTCCAGTTACTCT 1740
Db 1864 AAGAGTCTTCACTTCTTCTGCTGACCATCTTCAAGAGAGAGTCCAGTTACTCT 1923
Qy 1741 GCTCAGCTGACGACGCTGCTCACTTGGCGAGACCAAGTCTATCTGCTGCGAGGCG 1800
Db 1924 GCTCAGCTGACGACGCTGCTCACTTGGCGAGACCAAGTCTATCTGCTGCGAGGCG 1983
Qy 1801 CAGAAAGCATGACAGAGCTGACTGCGCGGAGCTGGCATGAAGAGAGCCCTTTGAA 1860

Db 1984 CAGAAAGCATGACAGAGCTGACTGCTGCGCGGAGCTGGCATGAAGAGAGCCCTTTGAA 2043
Qy 1861 AAGCAGTTTAAACGAGAAAGCTGCCAAATGAAATTTGAGAGAGCATGTCAGAGAAC 1920
Db 2044 AAGCAGTTTAAACGAGAAAGCTGCCAAATGAAATTTGAGAGAGCATGTCAGAGAAC 2103
Qy 1921 AGGTCAACGGGAAGAGCTGGGGAAGTGGGCACTGCTGACTTTTTCGGGACGATGAA 1980
Db 2104 AGGTCAACGGGAAGAGCTGGGGAAGTGGGCACTGCTGACTTTTTCGGGACGATGAA 2163
Qy 1981 ATCATTTAGGTCTCC 1995
Db 2164 ATCATTTAGGTCTCC 2178
RESULT 13
AX180875
LOCUS AX180875 2732 bp DNA linear PAT 06-AUG-2001
DEFINITION Sequence 2 from Patent WO0146394.
ACCESSION AX180875
VERSION AX180875.1 GI:15132703
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Plowman, G.D., Martinez, R., Whyte, D., Manning, G., Sudarsanam, S.,
Hill, R.J. and Planagan, P.
TITLE Mammalian protein phosphatases
JOURNAL Patent: WO 0146394-A 2 28-JUN-2001;
Sugen, Inc. (US)
FEATURES
source location/Qualifiers
1..2732
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
ORIGIN
Query Match 92.3%; Score 1842; DB 6; Length 2732;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1992; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 ATGGCCCATGAGATGATTTGGAACCTCAATTTGATGAGAGTGGGCTGCTGGAA 60
Db 538 ATGGCCCATGAGATGATTTGGAACCTCAATTTGATGAGAGTGGGCTGCTGGAA 597
Qy 61 AGTGAACGGAAGAGTGTCTGCTTAATTGATAGCCGACATTTGGAATCAATATCC 120
Db 598 AGTGAACGGAAGAGTGTCTGCTTAATTGATAGCCGACATTTGGAATCAATATCC 657
Qy 121 CACATTTTGAAGACCTTAATATCACTGCTCAAGCTTATGAAGAGAGTTCACAG 180
Db 658 CACATTTTGAAGACCTTAATATCACTGCTCAAGCTTATGAAGAGAGTTCACAG 717
Qy 181 GACAAAGTGTAAATTCAGAGCTCATCCAGCATTCAGGAAACATTAAGTTGACATTGAT 240
Db 718 GACAAAGTGTAAATTCAGAGCTCATCCAGCATTCAGGAAACATTAAGTTGACATTGAT 777
Qy 241 TGCAGTCAGAAAGTGTAGTTAGATTCAGTCAAGCTCCCAAGATGTTGCTCTCTTCA 300
Db 778 TGCAGTCAGAAAGTGTAGTTAGATTCAGTCAAGCTCCCAAGATGTTGCTCTCTTCA 837
Qy 301 GACTGTTTCTCACTGTAATCTCGGGTAACTGAGAAAGCTTCACTCTGTTCACTG 360
Db 838 GACTGTTTCTCACTGTAATCTCGGGTAACTGAGAAAGCTTCACTCTGTTCACTG 897
Qy 361 CTTCAGAGTGGTGGTGGTCTGCTGCTGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
Db 898 CTTCAGAGTGGTGGTGGTCTGCTGCTGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 957
Qy 421 ACTCTAGTCCCTTACTGCACTTCTGAGCTTCTTACTCTGTTGCAACATTTGGGCAACC 480

Db ACTGATCCCTACCTGATCTTCTCAGCCTTGCTTACCTGTGCAAACTTGGGCAACC 1017
QY 481 CGAATCTTCCCAATCTTTATCTTGGCTGCAAGAGATCTCTCAACAGAGCTGATA 540
Db 1018 CGAATCTTCCCAATCTTTATCTTGGCTGCAAGAGATCTCTCAACAGAGCTGATA 1077
QY 541 CAGCAGATGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
Db 1078 CAGCAGATGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1137
QY 601 ATCCCGAGTCTGATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660
Db 1138 ATCCCGAGTCTGATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1197
QY 661 CCGTGGTGGCAAACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 720
Db 1198 CCGTGGTGGCAAACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1257
QY 721 CTAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
Db 1258 CTAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1317
QY 781 AAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
Db 1318 AAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1377
QY 841 ATATCTCCAACTTCAATTTTCTGGGCAACTCTCTGATGATGATGATGATGATGAT 900
Db 1378 ATATCTCCAACTTCAATTTTCTGGGCAACTCTCTGATGATGATGATGATGATGAT 1437
QY 901 CAGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 960
Db 1438 CAGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1497
QY 961 GAACTGTCTCTGCTCTCAGAGGATGATGATGATGATGATGATGATGATGATGAT 1020
Db 1498 GAACTGTCTCTGCTCTCAGAGGATGATGATGATGATGATGATGATGATGATGAT 1557
QY 1021 TGTGCGAATCTGCTCTCAGAGGATGATGATGATGATGATGATGATGATGATGAT 1080
Db 1558 TGTGCGAATCTGCTCTCAGAGGATGATGATGATGATGATGATGATGATGATGAT 1617
QY 1081 CCCAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140
Db 1618 CCCAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1677
QY 1141 AGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1200
Db 1678 AGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1737
QY 1201 TCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1260
Db 1738 TCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1797
QY 1261 TCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320
Db 1798 TCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1857
QY 1321 AAGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1380
Db 1858 AAGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1917
QY 1381 GATTAAG 1440
Db 1918 GATTAAG 1977
QY 1441 AGCAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1500
Db 1978 AGCAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2037
QY 1501 TCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1560
Db 2038 TCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2097

QY 1561 CTTTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1620
Db 2098 CTTTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2157
QY 1621 TCGGATATCTTGGGCCCCCAGAGCTCTACCCCTTCTGACAGCAGCAGCTGATTTTGGC 1680
Db 2158 TCGGATATCTTGGGCCCCCAGAGCTCTACCCCTTCTGACAGCAGCAGCTGATTTTGGC 2217
QY 1681 AAGAGTCTCAACTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1740
Db 2218 AAGAGTCTCAACTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2277
QY 1741 GCTTACAGTGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1800
Db 2278 GCTTACAGTGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2337
QY 1801 CAGAGCCAAAGTGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1860
Db 2338 CAGAGCCAAAGTGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2397
QY 1861 AAGCAGTTAAACGAGAACTGCTGCAATGATGATGATGATGATGATGATGATGAT 1920
Db 2398 AAGCAGTTAAACGAGAACTGCTGCAATGATGATGATGATGATGATGATGATGAT 2457
QY 1921 AGGTACCGGAAAGAGCTGGGAAAGTGGGCAAGTCACTTTTGGGCAAGCAGGAA 1980
Db 2458 AGGTACCGGAAAGAGCTGGGAAAGTGGGCAAGTCACTTTTGGGCAAGCAGGAA 2517
QY 1981 ATCATTGAGTCTCC 1995
Db 2518 ATCATTGAGTCTCC 2532

RESULT 14
AX405700
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
CDS

1
Tang, Y. T., Liu, C., Zhou, P., Asundi, V., Zhang, J., Zhao, Q. A., Ren, F.,
Xue, A. J., Yang, Y., Wehrman, T., and Drmanac, R. T.
Novel nucleic acids and polypeptides
Patent: WO 0222660-A 115 21-MAR-2002;
HYSO, INC. (US)
Location/Qualifiers
1..3104
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
159..2156
/note="unassigned protein product"
/codon_start=1
/protein_id="CAD34807.1"
/db_xref="GI:21438840"
/db_xref="EMBL:CAD34807.1"
/translation="MAHEMIGQIVTERVALLSEGTENVLLIDSRPFVEYNTSHILE
AININCSKIMKRLQODKVLITELIOHSAKHVDIDCSQKVVYDQSDQVVASLSBDC
FLTVLLGKLEKSFNSVHLLAGFAFSEFPGICGSKFLVPTCISOPCLPVAINGPT
RLPVLVYGCORSDVUNKELMOONGYGVNANSTCPKPFIPESHPLRPVNDSECEK
ILPMDKSDTDFEAKASNGCTVYACLAGISASATILAIYIMKRDMSIDEXARPYKE
KPLTSPENFVGLUDYEKIKNGTGSAPSKULHLERPNBEPVAVSFGQKSE
TRSPCADSATSEBAAGRPVAPVSPVQPSILBDSPLVOLSGHLBADRLSD
SNKLRFSFIDIKSVYSASMAASLHGFSSSDALBYKFTLLDGTNLCOFSYVQES
LSBQPEPSPDKBASIPKLIOTARPDSQSKRLHSVRSSGTQORSLISPLHRSQS
VRDNYHTSFLRGLSTSOOHLTSAGLIGKMSHSDILAPOTSPLISLSMVPATSEHF
YASALITGGSASYSVSCQLPTCGDQVYSVRROKPSRDRASRSRSHSESPKQPK

ORIGIN RSCQMFGEISIMSENRSBELKIVGSSSFSGSMIEIIVS"

Query Match 92.3%; Score 1842; DB 6; Length 3104;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1992; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY 1 ATGGCCCATGAGATGATGGAATCTCAATTTGTAAGAGAGTGGTGGCTGCTGGAA 60
DB 159 ATGGCCCATGAGATGATGGAATCTCAATTTGTAAGAGAGTGGTGGCTGCTGGAA 218
QY 61 AGTGAACGGAAAAAGTGTCTAATTTGATAGCCGGCCATTTGTGAAATACATATACC 120
DB 219 AGTGAACGGAAAAAGTGTCTAATTTGATAGCCGGCCATTTGTGAAATACATATACC 278
QY 121 CACATTTTGGAAACCATTATATCAACTGTCTCAAGCTTATGAAAGGAAAGTTGCAACG 180
DB 279 CACATTTTGGAAACCATTATATCAACTGTCTCAAGCTTATGAAAGGAAAGTTGCAACG 338
QY 181 GACAAAGTGTATTTACAGAGCTCATCCGACATTGACGGAACATTAAGGTTGACATTGAT 240
DB 339 GACAAAGTGTATTTACAGAGCTCATCCGACATTGACGGAACATTAAGGTTGACATTGAT 398
QY 241 TGCAGTCAGAAAGTTGATTTACGATCAAAAGCTCCCAAGATGTTGCTCTCTCTCA 300
DB 399 TGCAGTCAGAAAGTTGATTTACGATCAAAAGCTCCCAAGATGTTGCTCTCTCTCA 458
QY 301 GACTGTTTTCTCACTGTACTTCTGGGTAACTGGAGAAAGCTTCAACTCTGTTCACTG 360
DB 459 GACTGTTTTCTCACTGTACTTCTGGGTAACTGGAGAAAGCTTCAACTCTGTTCACTG 518
QY 361 CTTCGAGTGGGTTGCTGAGTCTCTGCTGTTTCCTGGCTCTGTGAAGGAAATCC 420
DB 519 CTTCGAGTGGGTTGCTGAGTCTCTGCTGTTTCCTGGCTCTGTGTGAAGGAAATCC 578
QY 421 ACTCTAGTCCCTACCTGCAATTTCTCAGCCTTGCTTACCTGTGCGCAAACTGGGCCAAC 480
DB 579 ACTCTAGTCCCTACCTGCAATTTCTCAGCCTTGCTTACCTGTGCGCAAACTGGGCCAAC 638
QY 481 CGAATTTCTCCCAATCTTTATCTTGGGTGCGCAGCGAGATGCTCTCAACAGAGCTGATA 540
DB 639 CGAATTTCTCCCAATCTTTATCTTGGGTGCGCAGCGAGATGCTCTCAACAGAGCTGATA 698
QY 541 CAGCAGATGGAGTTGATGTTAATGACAGGTATACCTGTGCCAAAGCCTGACTT 600
DB 699 CAGCAGATGGAGTTGATGTTAATGACAGGTATACCTGTGCCAAAGCCTGACTT 758
QY 601 ATCCCGAGTCTCAATTTCTGCGGTGTGCTGTGATGACAGCTTTTGTGAAAAATTTTG 660
DB 759 ATCCCGAGTCTCAATTTCTGCGGTGTGCTGTGATGACAGCTTTTGTGAAAAATTTTG 818
QY 661 CCGTGTGGCAAAATCAGTATGATTGATTGAGAAAAGCAAAAGCCTCCATATGATGTT 720
DB 819 CCGTGTGTGGCAAAATCAGTATGATTGATTGAGAAAAGCAAAAGCCTCCATATGATGTT 878
QY 721 CTAGTCACTGTTTACCTGGGATCTCCGCGTCCGACCATGCTATGCTTACATCATG 780
DB 879 CTAGTCACTGTTTACCTGGGATCTCCGCGTCCGACCATGCTTACATCATG 938
QY 781 AAGAGATGAGACATGCTTTAGATGAAGCTTACAGATTTGTGAAAAAAAAGACCTACT 840
DB 939 AAGAGATGAGACATGCTTTAGATGAAGCTTACAGATTTGTGAAAAAAAAGACCTACT 998
QY 841 ATATCTCCAAATCTCAATTTTCTGGGCCAATCTCTGACTATGAGAAAGATTAAAGAC 900
DB 999 ATATCTCCAAATCTCAATTTTCTGGGCCAATCTCTGACTATGAGAAAGATTAAAGAC 1058
QY 901 CAGACTGAGCATCAGGGCCAAAGAGCAAACTCAAGCTGTGCACTGAGAGAGCCAAAT 960
DB 1059 CAGACTGAGCATCAGGGCCAAAGAGCAAACTCAAGCTGTGCACTGAGAGAGCCAAAT 1118
QY 961 GAACCTGTCCCTGCTCTCAGAGGTGACAGAAAAAGCAGAGCCCTCACTGCTCACC 1020
```

```
DB 1119 GAACCTGTCCCTGCTCTCAGAGGTGACAGAAAAAGCAGAGCCCTCACTGCTCACC 1178
QY 1021 TGTGCGGACTGTGCTACTCTCAGAGGACAGAGCAAAAGCCCTGCAATCCGCGCAGCG 1080
DB 1179 TGTGCGGACTGTGCTACTCTCAGAGGACAGAGCAAAAGCCCTGCAATCCGCGCAGCG 1238
QY 1081 CCCAGCGGCCAGGCTGAGCGCTGCTGTGTTAGAGAGCAGCCCGCTGTACAGGCGCTC 1140
DB 1239 CCCAGCGGCCAGGCTGAGCGCTGCTGTGTTAGAGAGCAGCCCGCTGTACAGGCGCTC 1298
QY 1141 AGTGGGCTTGACCTGTCTGCGACAGGCTTGAAAGCAGCAATTAAGCTCAAGCTTCTTC 1200
DB 1289 AGTGGGCTTGACCTGTCTGCGACAGGCTTGAAAGCAGCAATTAAGCTCAAGCTTCTTC 1358
QY 1201 TCTCTGATATCAATTCAGTTTCAATTACGACACATGAGAGCATCTCTACATGCTTC 1260
DB 1359 TCTCTGATATCAATTCAGTTTCAATTACGACACATGAGAGCATCTCTACATGCTTC 1418
QY 1261 TCTCTATGAGAAAGATGCTTTGGAATACAAACCTTCCACTTACCTGGAATGGACCAAC 1320
DB 1419 TCTCTATGAGAAAGATGCTTTGGAATACAAACCTTCCACTTACCTGGAATGGACCAAC 1478
QY 1321 AAGCTATGCGAGTTCTCCCTGTTCAAGAACTATCTGAGAGCAGATCTCCGAAAACAATCT 1380
DB 1479 AAGCTATGCGAGTTCTCCCTGTTCAAGAACTATCTGAGAGCAGATCTCCGAAAACAATCT 1538
QY 1381 GATTAAGAGAAAGCAGAGATCCCAAGAAAGCTGCGACACCGCAGGCTTTCAGAGCGCAG 1440
DB 1539 GATTAAGAGAAAGCAGAGATCCCAAGAAAGCTGCGACACCGCAGGCTTTCAGAGCGCAG 1598
QY 1441 AGCAAGCATATGCTATCGGTGAGAACCCAGCAGATGCGACCGCCAGAGGTCCCTTTTA 1500
DB 1599 AGCAAGCATATGCTATCGGTGAGAACCCAGCAGATGCGACCGCCAGAGGTCCCTTTTA 1658
QY 1501 TCTCCATGTCATCGAAGTGGAGCGTGGAGCAATTAACAACAAGCTTCTTTTCGCG 1560
DB 1659 TCTCCATGTCATCGAAGTGGAGCGTGGAGCAATTAACAACAAGCTTCTTTTCGCG 1718
QY 1561 CTTTCACACAGCAGACACACCTCAAGAAAGTGTGCGCTGAGGCTTAAGGGCTGGAC 1620
DB 1719 CTTTCACACAGCAGACACACCTCAAGAAAGTGTGCGCTGAGGCTTAAGGGCTGGAC 1778
QY 1621 TCGATATCTTGGCCCCCAGACCTCTACCCCTTCCCTGACCAAGCAGCTGATTTTGGC 1680
DB 1779 TCGATATCTTGGCCCCCAGACCTCTACCCCTTCCCTGACCAAGCAGCTGATTTTGGC 1838
QY 1681 ACAGAGTCTCACTTCTACTCTGCTGAGCATCTACGAGGAGCAGTCCAGTTACTCT 1740
DB 1839 ACAGAGTCTCACTTCTACTCTGCTGAGCATCTACGAGGAGCAGTCCAGTTACTCT 1898
QY 1741 GCTTACAGTGAAGCAGCTGCGCACTTGGCGGAGCAAGTCTATTGTGCGCAGGGCG 1800
DB 1899 GCTTACAGTGAAGCAGCTGCGCACTTGGCGGAGCAAGTCTATTGTGCGCAGGGCG 1958
QY 1801 CAGAAAGCAAGTGAAGAGCTGACTCGCGCGGAGACTGGCAATGAAGAGCCCTTTGAA 1860
DB 1959 CAGAAAGCAAGTGAAGAGCTGACTCGCGCGGAGACTGGCAATGAAGAGCCCTTTGAA 2018
QY 1861 AAGCAGTTTAAACGAGAAAGCTGCCAAATGGAATTTGAGAGAGCATCATGTCAAGAAC 1920
DB 2019 AAGCAGTTTAAACGAGAAAGCTGCCAAATGGAATTTGAGAGAGCATCATGTCAAGAAC 2078
QY 1921 AGGTACCGGGAAGACTGGGGAAGTGGGAGTCAAGTCTAGCTTTTGGGGCAGATGAA 1980
DB 2079 AGGTACCGGGAAGACTGGGGAAGTGGGAGTCAAGTCTAGCTTTTGGGGCAGATGAA 2138
QY 1981 ATCATTTAGAGTCTCC 1995
DB 2139 ATCATTTAGAGTCTCC 2153
```

RESULT 15
AX713989

Db 1796 GCCTACAGCTGCAGCCAGGCTGCCACTGTGGAGAGCAAGTCTATTCTGTGGCGCAGCGG 1855

QY 1801 CAGAACCCAGATACAGAGTGACTCGCGCGGAGCTGGCATGAAGAGCCCTTTGAA 1860

Db 1856 CAGAACCCAGATACAGAGTGACTCGCGCGGAGCTGGCATGAAGAGCCCTTTGAA 1915

QY 1861 AAGCATTTAAACGACGAGCTCCAAATGGAATTTGAGAGAGCATCATGTGCAGAAC 1920

Db 1916 AAGCATTTAAACGACGAGAGCTCCAAATGGAATTTGAGAGAGCATCATGTGCAGAAC 1975

QY 1921 AGGTCAAGGAGAGAGCTGGGAGAGTGGGAGTCACTTACCTTTTGGGAGCATGAA 1980

Db 1976 AGGTCAAGGAGAGAGCTGGGAGAGTGGGAGTCACTTACCTTTTGGGAGCATGAA 2035

QY 1981 ATCATTTAGTCTCC 1995

Db 2036 ATCATTTAGTCTCC 2050

RESULT 16
AK055973

LOCUS 2102 bp mRNA linear PRI 01-AUG-2002

DEFINITION Homo sapiens cDNA FLJ31411 fis, clone NT2NE2000214, moderately similar to DUAL SPECIFICITY PROTEIN PHOSPHATASE 8 (EC 3.1.3.48).

ACCESSION AK055973

VERSION AK055973.1 GI:16550835

KEYWORDS oligo capping; fis (full insert sequence).

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 Tashiro, H., Yamazaki, M., Watanabe, K., Kumagai, A., Itakura, S., Fukushima, Y., Fujimori, Y., Komiyama, M., Sugiyama, T., Ito, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Matsushima, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahata, K., Masuh, Y., Nagai, K. and Isogai, T.

TITLE NED0 human cDNA sequencing project

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 2102)

AUTHORS Isogai, T., Otsuki, T. and Sugiyama, T.

TITLE Direct Submission

JOURNAL Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail:genom@csb.hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)

COMMENT NED0 human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: RAB and HRI.

FEATURES

SOURCE

1. .2102

Location/Qualifiers

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="NT2NE2000214"

/cell_line="NT2"

/cell_type="teratocarcinoma"

/clone_lib="NT2NE2"

/note="Cloning vector: pME18FLJ3-mRNA from NT2 neuron after the differentiation of NT2 neuronal precursor cells."

56..2053

/note="unnamed protein product"

/codon_start=1

/protein_id="BAB71060.1"

/db_xref="GI:16550835"

/translation="MAHEMIGQVYTERVALLESGETEKVLIDSRPEVNTSHILE
AINVSGKLMKRLQDDVLTLEI QHSAKRIIDGQKVVYDSSDVAASDC
FLTVLGLKERSFNSVHLLAGFAFSPGLCGKSLVPTCTISQPLPAVANGPT
LILPMLYIGQEDVNLKELMOONGIGVYLNASNTCPKDPFIPEHFLLVPVNSCEK
LILPMDKVDDELEKAKNSGCVLNGVIGSRSATIAIAYIMKRMDSIDEAYRFEKE
KRPTISPNFNLGOLLIDYEKKIKNOTGASGPKSKLHLLEKPNPVAVSQGGKSE
TLPSPCADSATSEARQPVHPASVPSVPSVOPSLDPSLPYVALSGLHLSADLEL
SNKLRSPSLDIKSVSYSAASAISLHGSSSEDALEYKRPETTLDTGNTKLCPSVOE
LSEQTPETSPPKEASIPKIQTAAPSQSRRLHSVRSSTGTRSLSPHRS
VEDNHTSFLFGLSTSOQLTKSAGLGLKMHSDILAPQTSPLTSSWYATSESHF
YSASAIYGSASYSVNSCSOLPTGQDYQVSVRRQKPDRAADRSRMHEBSFEKQFK
RRSCMERGESIMSENRSREIKGVKSQSSPSFSMEIIEVS"

Query Match 89.8%; Score 1791; DB 9; Length 2102;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1991; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATGGCCCATGAGATGATTTGAACTCAATTTGTTACTGAGAGTGGTGGCTCTGCTGAA 60

Db 56 ATGGCCCATGAGATGATTTGAACTCAATTTGTTACTGAGAGTGGTGGCTCTGCTGAA 115

QY 61 AGTGAACGGAAGAAAGTCTGCTAATTGATAGCCGCGCATTTTGGAATACAAATACATCC 120

Db 116 AGTGAACGGAAGAAAGTCTGCTAATTGATAGCCGCGCATTTTGGAATACAAATACATCC 175

QY 121 CACATTTTGAAGACCATTAATATCACTGCTCCAAAGCTTTATGAAACGGAAGTTGCAACG 180

Db 176 CACATTTTGAAGACCATTAATATCACTGCTCCAAAGCTTTATGAAACGGAAGTTGCAACG 235

QY 181 GACAAAGTTAATATACAGAGTCAATCCAGATTCAGGAAACATPAAGTTGACATTGAT 240

Db 236 GACAAAGTTAATATACAGAGTCAATCCAGATTCAGGAAACATPAAGTTGACATTGAT 295

QY 241 TGCAGTCAGAAAGTTGATGATTAAGATCAAAAGCTCCAAAGATGTTGCTCTCTCTTCA 300

Db 296 TGCAGTCAGAAAGTTGATGATTAAGATCAAAAGCTCCAAAGATGTTGCTCTCTCTTCA 355

QY 301 GACTGTTTTCACACTGATCTCTGATTAACATCGAAGAGACTTCAATCTGTTCACTG 360

Db 356 GACTGTTTTCACACTGATCTCTGATTAACATCGAAGAGACTTCAATCTGTTCACTG 415

QY 361 CTGCAAGTGGATTTGCTGAGATTCCTGCTGTTTCCCTGSCCTCTGGAAGAAATACC 420

Db 416 CTGCAAGTGGATTTGCTGAGATTCCTGCTGTTTCCCTGSCCTCTGGAAGAAATACC 475

QY 421 ACTCTAGTCCCTACCTGATTTCTCAGCCTTCTTACCTGTTGCAACATTTGGGCAACC 480

Db 476 ACTCTAGTCCCTACCTGATTTCTCAGCCTTCTTACCTGTTGCAACATTTGGGCAACC 535

QY 481 CGAATTTCTCCCAATCTTTAATCTTGGCTGCGCAGAGATGTCCTCAACAGAGCTGATA 540

Db 536 CTAATTTCTCCCAATCTTTAATCTTGGCTGCGCAGAGATGTCCTCAACAGAGCTGATA 595

QY 541 CAGCAGAAATGGAGTTGGTATGTTAAATGCGAGCTATACCTGCAAAAGCTGACCTT 600

Db 596 CAGCAGAAATGGAGTTGGTATGTTAAATGCGAGCTATACCTGCAAAAGCTGACCTT 655

QY 601 ATCCCGAGTCAATTTCTGCGCTGCTGCTGATGACAGCTTTTGGAGAAATTTTG 660

Db 656 ATCCCGAGTCAATTTCTGCGCTGCTGCTGATGACAGCTTTTGGAGAAATTTTG 715

QY 661 CCGGTGTTGGAACAATAGATGATTTCAATGAGAAAGCAAAAGCTCAATGATGTT 720

Db 716 CCGGTGTTGGAACAATAGATGATTTCAATGAGAAAGCAAAAGCTCAATGATGTT 775

QY 721 CTAAGCAGCTGTTAGCGGGATCCCGGCTCGGCAACATGCTATGCGCTATCATATG 780

Db 776 CTAAGCAGCTGTTAGCGGGATCCCGGCTCGGCAACATGCTATGCGCTATCATATG 835

QY 781 AAGAGATGACATGCTTTAGATGAAGCTTACAGATTTGTAAGAAAGAAAGCTTACT 840

Db 836 AAGAGATGAGATGCTTTAGATGAGCTTACAGATTTGTGAAAGAAAAAGACTACT 895
 QY 841 ATATCTCAAACTTCAATTTTCTGGGCAACTCTGGACTATGAGAAAGATTAAAGAC 900
 Db 896 ATATCTCAAACTTCAATTTTCTGGGCAACTCTGGACTATGAGAAAGATTAAAGAC 955
 QY 901 CAGACTGAGACATCAGGGCCAAAGAGCAACTCAAGCTGTGCACTGTGAAAGCCAAAT 960
 Db 956 CAGACTGAGACATCAGGGCCAAAGAGCAACTCAAGCTGTGCACTGTGAAAGCCAAAT 1015
 QY 961 GAACTGTCTCTGTCTCTAGAGGTTGACAGAAAAAGCAGACGCCCTCACTCCACC 1020
 Db 1016 GAACTGTCTCTGTCTCTAGAGGTTGACAGAAAAAGCAGACGCCCTCACTCCACC 1075
 QY 1021 TGTGCGCACTCTGTCTCTAGAGGTTGACAGAAAAAGCAGACGCCCTCACTCCACC 1080
 Db 1076 TGTGCGCACTCTGTCTCTAGAGGTTGACAGAAAAAGCAGACGCCCTCACTCCACC 1135
 QY 1081 CCCAGCGTGCACGCTGACGCGTGTGTAGAGGACAGCGCGTGTGTAGAGGCGCTC 1140
 Db 1136 CCCAGCGTGCACGCGTGTGTAGAGGACAGCGCGTGTGTAGAGGCGCTC 1195
 QY 1141 AGTGGCTGCACTGTCTCCGACAGAGCTGGAAGACAGCAATTAAGCTCAAGCTTCTTC 1200
 Db 1196 AGTGGCTGCACTGTCTCCGACAGAGCTGGAAGACAGCAATTAAGCTCAAGCTTCTTC 1255
 QY 1201 TCTCTGATATCAATCAAGTTTCAATTCAGAGGACATGCGACATCTTCAATGCTTC 1260
 Db 1256 TCTCTGATATCAATCAAGTTTCAATTCAGAGGACATGCGACATCTTCAATGCTTC 1315
 QY 1261 TCTCTATCAGAAAGATGCTTTGGAATACTAACAACCTTCTCACTGTGAGGACCAAC 1320
 Db 1316 TCTCTATCAGAAAGATGCTTTGGAATACTAACAACCTTCTCACTGTGAGGACCAAC 1375
 QY 1321 AAGCTATGCCAGTTCTTCCCTGTTCAAGAACTATCGAGAGACACTCCGAAACCAAGTCT 1380
 Db 1376 AAGCTATGCCAGTTCTTCCCTGTTCAAGAACTATCGAGAGACACTCCGAAACCAAGTCT 1435
 QY 1381 GATTAAGAGGAGAACCGACATCCCGCAAGACCTCAGACCGCGCTTCAAGACCGAC 1440
 Db 1436 GATTAAGAGGAGAACCGACATCCCGCAAGACCTCAGACCGCGCTTCAAGACCGAC 1495
 QY 1441 AGCAAGCATTTGATTCGATTCAGAAACAGCAGAGTGGACCGCCAGAGAGTCCCTTTTA 1500
 Db 1496 AGCAAGCATTTGATTCGATTCAGAAACAGCAGAGTGGACCGCCAGAGAGTCCCTTTTA 1555
 QY 1501 TCTCACTGATGAAAGTGGAGCGTGGAGACATTAACAACAGCTTCTTTTCGAG 1560
 Db 1556 TCTCACTGATGAAAGTGGAGCGTGGAGACATTAACAACAGCTTCTTTTCGAG 1615
 QY 1561 CTTTCCACGACGACGACGACCTCAAGAACTCTGCTGGCTTGAAGGCTTGGCAC 1620
 Db 1616 CTTTCCACGACGACGACGACCTCAAGAACTCTGCTGGCTTGAAGGCTTGGCAC 1675
 QY 1621 TCGGATATCTTTGGCCCCCAGACCTCAACCCCTTCCAGACGAGCTGTATTTTGGC 1680
 Db 1676 TCGGATATCTTTGGCCCCCAGACCTCAACCCCTTCCAGACGAGCTGTATTTTGGC 1735
 QY 1681 ACAGAGTCTCTACACTTCTACTCTGCTCAGCCATCTACGAGAGCAGTGGCACTACT 1740
 Db 1736 ACAGAGTCTCTACACTTCTACTCTGCTCAGCCATCTACGAGAGCAGTGGCACTACT 1795
 QY 1741 GCTTACAGCTGACGACGACCTTGGCCACTTGGGAGACCAAGTCTTATTTGTGGCAGGCG 1800
 Db 1796 GCTTACAGCTGACGACGACCTTGGCCACTTGGGAGACCAAGTCTTATTTGTGGCAGGCG 1855
 QY 1801 CAGAGCAGATGACAGAGCTGACTCGCGCGAGAGTGGCATGAAGAGAGCCCTTTGAA 1860
 Db 1856 CAGAGCAGATGACAGAGCTGACTCGCGCGAGAGTGGCATGAAGAGAGCCCTTTGAA 1915
 QY 1861 AAGCAGTTTAAACGACAGACCTTGCCTCAATATTTGAGAGAGCATCATGTACAGAAC 1920
 Db 1916 AAGCAGTTTAAACGACAGACCTTGCCTCAATATTTGAGAGAGCATCATGTACAGAAC 1975

QY 1921 AGCTACGGGAGAGAGCTGGGGAAAGTGGCAGTCACTAGCTTTTCTGGGACAGATGAA 1980
 Db 1976 AGCTACGGGAGAGAGCTGGGGAAAGTGGCAGTCACTAGCTTTTCTGGGACAGATGAA 2035
 QY 1981 ATCATTGAGGTCTCC 1995
 Db 2036 ATCATTGAGGTCTCC 2050
 RESULT 17
 AX482372 5111 bp DNA linear PAT 16-AUG-2002
 LOCUS AX482372
 DEFINITION Sequence 41 from Patent WO02057460.
 ACCESSION AX482372
 VERSION AX482372.1 GI:22316945
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1
 AUTHORS Todderud, C.G., Bol, D., Finger, J., Lee, L., Nelson, T., Schieven, G.,
 Suchard, S., Banas, D., Bassolino, D., Feder, J., Kyatek, S.,
 Mcatee, P., Mintier, G., Siemers, N., Jackson, D.G., and Ramanathan, C.
 TITLE Polynucleotides encoding human phosphatases
 JOURNAL Patent: WO 02057460-A 41 25-JUL-2002;
 SQUIBB BRISTOL MYERS CO (US)
 FEATURES
 source 1..5111
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 ORIGIN
 Query Match 89.8%; Score 1791; DB 6; Length 5111;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1991; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 ATGGCCCATGATGATGTAAGTAAGTCAATTTGTTACTGAGAGTTGGTGTCTGCTGAA 60
 Db 470 ATGGCCCATGATGATGTAAGTAAGTCAATTTGTTACTGAGAGTTGGTGTCTGCTGAA 529
 QY 61 AGTGAAGGGAAGAAAGTGTGCTAATTGATAGCGGSCATTGTGGAATTAACATATCC 120
 Db 530 AGTGAAGGGAAGAAAGTGTGCTAATTGATAGCGGSCATTGTGGAATTAACATATCC 589
 QY 121 CACATTTTGAAGCCATTATATCACTGCTCCAGCTTATGAAAGGAGTTGCAACAG 180
 Db 590 CACATTTTGAAGCCATTATATCACTGCTCCAGCTTATGAAAGGAGTTGCAACAG 649
 QY 181 GACAAAGTTTAATTACAGAGCTCATCCAGACTTACGGAACATTAAGTTGACATTAT 240
 Db 650 GACAAAGTTTAATTACAGAGCTCATCCAGACTTACGGAACATTAAGTTGACATTAT 709
 QY 241 TGGAGTCAGAAAGTTGATTAAGATCAAGTCCCAAGATGGTGGCTCTCTCTTA 300
 Db 710 TGGAGTCAGAAAGTTGATTAAGATCAAGTCCCAAGATGGTGGCTCTCTCTTA 769
 QY 301 GACTGTTTCTCACTGTAATTTCTGGGTAAGTGAAGAGAGTTCAACTCTGTTACCTG 360
 Db 770 GACTGTTTCTCACTGTAATTTCTGGGTAAGTGAAGAGAGTTCAACTCTGTTACCTG 829
 QY 361 CTTCAGAGTGGGTTTGTCTGAGTTCTCTGTTTTCCTGGCTCTGTGAAGAAATCC 420
 Db 830 CTTCAGAGTGGGTTTGTCTGAGTTCTCTGTTTTCCTGGCTCTGTGAAGAAATCC 889
 QY 421 ACTCTAGTCCCTACCTGATTTCTGAGCTTGTCTTACCTGTTGCCAAGATTGGGCCAAC 480
 Db 890 ACTCTAGTCCCTACCTGATTTCTGAGCTTGTCTTACCTGTTGCCAAGATTGGGCCAAC 949
 QY 481 CGAATTCCTCCCAATCTTTATCTTGGCTGCCAGGAGATGTCTCAACAGAGAGCTGATA 540

Db	950	CGAATCTTCCCAATCTTTATCTTGCGTCCAGGAGATGTCTCAACAAGAGCTATG	1009
Qy	541	CAGAGATATGGATTTGGTATGTGTAAATGCCAGCTATACCTGTCCAAAGCTGACTTT	600
Db	1010	CAGCAAAATGGATTTGGTATGTGTAAATGCCAGCAATACCTGTCCAAAGCTGACTTT	1069
Qy	601	ATCCCGAGTCTCAATTTCTGCGTGTGTGAATGACGCTTTTGTGAATAATTTTG	660
Db	1070	ATCCCGAGTCTCAATTTCTGCGTGTGTGAATGACGCTTTTGTGAATAATTTTG	1129
Qy	661	CCGTGTGTGAACAATCAGTATGATTTCAAGAAACCAAAAGCTTCAATGTGTGT	720
Db	1130	CCGTGTGTGAACAATCAGTATGATTTCAAGAAACCAAAAGCTTCAATGTGTGT	1189
Qy	721	CTAGTGCATGTTTAAAGCGGGATCTCCCGTCCGCGCAATCGCTATCGCTACATCAG	780
Db	1190	CTAGTGCATGTTTAAAGCGGGATCTCCCGTCCGCGCAATCGCTATCGCTACATCAG	1249
Qy	781	AAGAGATGACATGTCTTTAGATGAAGCTTACAGATTTGTGAAGAAAGAAAGCTTACT	840
Db	1250	AAGAGATGACATGTCTTTAGATGAAGCTTACAGATTTGTGAAGAAAGAAAGCTTACT	1309
Qy	841	ATATCTCCAACTTCAATTTTCTGGGCCAACTCCTGCACTATGAGAAGAAATTAAAGAC	900
Db	1310	ATATCTCCAACTTCAATTTTCTGGGCCAACTCCTGCACTATGAGAAGAAATTAAAGAC	1369
Qy	901	CAGACTGAGCATCAGGGCCAAAGAGCAACTCAGCTGTGACCTCGAGAGAGCCCAAT	960
Db	1370	CAGACTGAGCATCAGGGCCAAAGAGCAACTCAGCTGTGACCTCGAGAGAGCCCAAT	1429
Qy	961	GAACCTGTCTCTGCTGTCTCAGAGGTGACAGAAAAAGCAGAGCCCTCAGTCCACCC	1020
Db	1430	GAACCTGTCTCTGCTGTCTCAGAGGTGACAGAAAAAGCAGAGCCCTCAGTCCACCC	1489
Qy	1021	TTGGCCGACCTCTCTCCTCAGAGGACAGAGCAAAAGCCCTGATCCCGCAGCTGT	1080
Db	1490	TTGGCCGACCTCTCTCCTCAGAGGACAGAGCAAAAGCCCTGATCCCGCAGCTGT	1549
Qy	1081	CCGAGCGTGCACGCTGACGCGTGCCTGTGTAGAGACAGCCCGTGTACAGAGCGCTC	1140
Db	1550	CCGAGCGTGCACGCTGACGCGTGCCTGTGTAGAGACAGCCCGTGTACAGAGCGCTC	1609
Qy	1141	AGTGGCTGACCTGTCCGACAGAGCTGGAAGACAGCAATTAAGCTCAAGCGTTCCTTC	1200
Db	1610	AGTGGCTGACCTGTCCGACAGAGCTGGAAGACAGCAATTAAGCTCAAGCGTTCCTTC	1669
Qy	1201	TCTCTGATATCAAAATCAGTTTCATTTGACGACAGATGAGCATCTTATATGCTTC	1260
Db	1670	TCTCTGATATCAAAATCAGTTTCATTTGACGACAGATGAGCATCTTATATGCTTC	1729
Qy	1261	TCTCTATCAGAAAGATCTTTGAAATCTACAAACCTTCCACTACTGATGGAGCAAC	1320
Db	1730	TCTCTATCAGAAAGATCTTTGAAATCTACAAACCTTCCACTACTGATGGAGCAAC	1789
Qy	1321	AAGCTATGCAATTTCTCCCTGTTTGAAGAACTATGAGACAGACTCCGAAACAGTCT	1380
Db	1790	AAGCTATGCAATTTCTCCCTGTTTGAAGAACTATGAGACAGACTCCGAAACAGTCT	1849
Qy	1381	GATTAAGAGAGAGCCAGCATCCCAAGAGCTGACAGCCGAGCCCTTACAGAGCCAG	1440
Db	1850	GATTAAGAGAGAGCCAGCATCCCAAGAGCTGACAGCCGAGCCCTTACAGAGCCAG	1909
Qy	1441	AGCAAGCAATTTGCTGCTGAGAACAGACAGCAAGTGGACCGCCAGAGGTCCCTTTTA	1500
Db	1910	AGCAAGCAATTTGCTGCTGAGAACAGACAGCAAGTGGACCGCCAGAGGTCCCTTTTA	1969
Qy	1501	TCTCCAGTGCATCGAAGTGGAGCTGAGAGCAATTAACAACACAGCTTCTTTTGGGC	1560
Db	1970	TCTCCAGTGCATCGAAGTGGAGCTGAGAGCAATTAACAACACAGCTTCTTTTGGGC	2029
Qy	1561	CTTTCCACGAGCAGACACACTCAGAAATCTGCTGGCTTTAAAGGGCTGGCAC	1620
Db	2030	CTTTCCACGAGCAGACACACTCAGAAATCTGCTGGCTTTAAAGGGCTGGCAC	2089

Qy	1621	TCGATATCTTGAGCCGCCAGACCTTACCCCTTCTGACAGCAGAGCTGATTTTGGC	1680
Db	2090	TCGATATCTTGAGCCGCCAGACCTTACCCCTTCTGACAGCAGAGCTGATTTTGGC	2149
Qy	1681	ACAGAGTCTTCAACTTTCTTCTGCTGAGCATCTACAGAGAGTGGACGATTACTCT	1740
Db	2150	ACAGAGTCTTCAACTTTCTTCTGCTGAGCATCTACAGAGAGTGGACGATTACTCT	2209
Qy	1741	GCCTACAGTGCAGCAGCTGCGCACTTGCGGAGCAAGCTATTTCTGTGCGAGCGGG	1800
Db	2210	GCCTACAGTGCAGCAGCTGCGCACTTGCGGAGCAAGCTATTTCTGTGCGAGCGGG	2269
Qy	1801	CAGAACCAAGTACAGAGCTGACTCGCGCGAGCTGGCATGAGAGAGCCCTTTGAA	1860
Db	2270	CAGAACCAAGTACAGAGCTGACTCGCGCGAGCTGGCATGAGAGAGCCCTTTGAA	2329
Qy	1861	AAGCAGTTTAAACGAGAGAGCTGCCAATTGAATTTGAGAGAGCATATGTACAGAAC	1920
Db	2330	AAGCAGTTTAAACGAGAGAGCTGCCAATTGAATTTGAGAGAGCATATGTACAGAAC	2389
Qy	1921	AGGTACCGGAAAGAGCTGGGAAAGTGGGACGTCAGTCACTTTTGGGAGAGTGA	1980
Db	2390	AGGTACCGGAAAGAGCTGGGAAAGTGGGACGTCAGTCACTTTTGGGAGAGTGA	2449
Qy	1981	ATCATTTGAGTCTCC	1995
Db	2450	ATCATTTGAGTCTCC	2464

RESULT 18
 BC031643
 LOCUS
 DEFINITION
 Homo sapiens, clone IMAGE:5176724, mRNA, partial cde.
 ACCESSION
 BC031643
 VERSION
 BC031643.1 GI:21594972
 KEYWORDS

SOURCE
 ORGANISM
 Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 2807)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL

Direct Submission
 Submitted (06-JUN-2002) National Institutes of Health, Mammalian
 Gene Collection (MGC), Cancer Genomics Office, National Cancer
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
 USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk

Email: cgabs-remail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: Baylor College of Medicine Human Genome
 Sequencing Center

Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: amg@bcm.tmc.edu

Guarnarone, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
 Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
 Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
 through the I.M.A.G.E. Consortium/BLNI at: <http://image.llnl.gov>
 Series: IRAX Plate: 51 Row: n Column: 21
 This clone was selected for full length sequencing because it
 passed the following selection criteria: Hexamer frequency ORF
 analysis.

FEATURES
 source

1..2807
 /organism="Homo sapiens"
 /mol_type="mRNA"

/db_xref="taxon:9606"
/clone="IMAGE:5176724"
/tissue_type="Brain, lung, Testis, adult, pooled whole"
/clone_id="NH_MGC_115"
/lab_host="DH10B"
/note="Vector: pCMV-Sport6"
<1..1852
/codon_start=2
/product="Unknown (protein for IMAGE:5176724)"
/protein_id="AAH31643.1"
/db_xref="GI:21594973"
/translation="CSKLMKRLQODKVLTELIOHSAKHVIDPSOKVYVDDSSQ
DVASDDCLFVLGLKESFNSVHLAAGFAFSPGCEGSLVPTCTISPC
LPVANTGPRILPNLYIGCORVLANKEIMQONGIVLANSTCPKDPFPESHFLRV
PVNDSFEKILPMLDKSVDFIEKAKSNQCVLHCLAGISRSATIAIYIMKRMDSL
DEAVRFLVEKERPTISPNFNLGQLIDBEKIKNOTGASGPKSLKULHEKEPVPA
VSEGGQSETPPLSPPCSDATSEAAQORPHAPASVPSVPSLLEDEPLVQALGL
HLSADRLDESNKLRSFSLDKSVSASMAASLHGSSEDELEVYKSTLIDGINK
LCQSPVQELSEOTPEPSPDKEASIPKLOTPRSPDSOKRLSHVRSISGTAORSL
LSPLHSGSVEDNHTSFLRGLSTSOOHLTKSAGLIGKGMHSLIAPORTSPSTLSM
YPTASHSHFASAIYGGASYSAYSCSOLPTGDOYTVRRROKPSDRADSRSHWE
ESPFEKQFKRSQMEGESIMSENSREELGRVGSQSSFSGSEIIEVS"

CDS

ORIGIN

Query Match 85.0%; Score 1696; DB 9; Length 2807;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1846; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

147 CTGCTCCAGCTTATGAAGCGAAGTTGCAACAGCAAGTGTATTAACAGAGCTCAT 206
1 CTGCTCCAGCTTATGAAGCGAAGTTGCAACAGCAAGTGTATTAACAGAGCTCAT 60
207 CCAGCATTCAGCGAACAATAGTTGACATTGATGATGATGATGATGATGATGATGAT 266
61 CCAGCATTCAGCGAACAATAGTTGACATTGATGATGATGATGATGATGATGATGAT 120
267 TCAAGCTCCAGAGTGTGCTCTCTCTGATGATGATGATGATGATGATGATGATGAT 326
121 TCAAGCTCCAGAGTGTGCTCTCTCTGATGATGATGATGATGATGATGATGATGAT 180
327 TAACTGAGAGAGCTTCAACTCTGTTCACTGCTTGCAGAGTGGGTTGCTGAGTTCTC 386
181 TAACTGAGAGAGCTTCAACTCTGTTCACTGCTTGCAGAGTGGGTTGCTGAGTTCTC 240
387 TCGTTGTTCCCTGCGCTCTGTGAAGAAATCCACTCTAGTCCCTTACCTGCAATTTCTCA 446
241 TCGTTGTTCCCTGCGCTCTGTGAAGAAATCCACTCTAGTCCCTTACCTGCAATTTCTCA 300
447 GCCTTGCTTACCTGTTGCCAATTTGGGCCAACCCGGAATCTTCCGAATCTTATATCTTGG 506
301 GCCTTGCTTACCTGTTGCCAATTTGGGCCAACCCGGAATCTTCCGAATCTTATATCTTGG 360
507 CTGCGACGAGATGTCTCTCAACAGAGCTGATACAGAGATGGAGTTGGTATGTTGT 566
361 CTGCGACGAGATGTCTCTCAACAGAGCTGATACAGAGATGGAGTTGGTATGTTGT 420
567 AAATGCGAGCTATACCTGTCCAAAGCCTGACTTTATCCCGAGTCTCAATTTCTGCGGT 626
421 AAATGCGAGCTATACCTGTCCAAAGCCTGACTTTATCCCGAGTCTCAATTTCTGCGGT 480
627 GCCTGTGAATGACAGCTTTTGTGAGAAATTTTGGCGTGGTGGACAAATCAAGTAGTTT 686
481 GCCTGTGAATGACAGCTTTTGTGAGAAATTTTGGCGTGGTGGACAAATCAAGTAGTTT 540
687 CATTGAGAAAGCAAAAGCCTTCAATGGATGTTGTTCTAGTGCACTGTTAGCTGGAGTCTC 746
541 CATTGAGAAAGCAAAAGCCTTCAATGGATGTTGTTCTAGTGCACTGTTAGCTGGAGTCTC 600
747 CCCTGCGGCCACCATGCTATCGCTTACATCATGAGAGATGGAGATGTTCTTTAGATGA 806
601 CCCTGCGGCCACCATGCTATCGCTTACATCATGAGAGATGGAGATGTTCTTTAGATGA 660
807 AGCTTACAGATTTGTGAGAAAGAAAGCACTACTATATCTTCCAAATTTCTCAATTTTCTGGG 866

661 AGCTTACAGATTTGTGAGAAAGAAAGCACTACTATATCTTCCAAATTTTCTGGG 720
867 CCAGCTCTCTGAGCTATGAGAGAAATTAAGAACCGAGTCTGAGATCTAGGCGCAAGAG 926
721 CCAGCTCTCTGAGCTATGAGAGAAATTAAGAACCGAGTCTGAGATCTAGGCGCAAGAG 780
927 CAAACTCAAGCTGCTGACCTGAGAGAGGCAATGAACTGTCCCTGCTGTCTCAGAGGG 986
781 CAACTCAAGCTGCTGACCTGAGAGAGGCAATGAACTGTCCCTGCTGTCTCAGAGGG 840
987 TGGACAGAAAGAGAGAGAGCCCTCAGTCACCTGTGCGCATCTGTCTCAGAGGC 1046
841 TGGACAGAAAGAGAGAGAGCCCTCAGTCACCTGTGCGCATCTGTCTCAGAGGC 900
1047 AGCAGAGCAAAAGCCCTGATATCCCGCCAGCTGCGCCAGCTGCGCCAGAGCTGCGCC 1106
901 AGCAGAGCAAAAGCCCTGATATCCCGCCAGCTGCGCCAGCTGCGCCAGAGCTGCGCC 960
1107 GCTGTTAGAGAGACAGCCCGCTGGTACAGCGGCTCAGTGGGCTGACCTGTCCGACAGAG 1166
961 GCTGTTAGAGAGAGAGCCCGCTGGTACAGCGGCTCAGTGGGCTGACCTGTCCGACAGAG 1020
1167 GCTGAGAGACAGCAATTAAGCTCAAGCGTCTCTCTCTGATATCAAAATCAATTTCTTCA 1226
1021 GCTGAGAGACAGCAATTAAGCTCAAGCGTCTCTCTCTGATATCAAAATCAATTTCTTCA 1080
1227 TTCAGCCAGATGAGAGATCTCTTACATGAGCTTCTCTTCAATCAAGAGTCTTGGAAAT 1286
1081 TTCAGCCAGATGAGAGATCTCTTACATGAGCTTCTCTTCAATCAAGAGTCTTGGAAAT 1140
1287 CTACAAACCTTCCACTACTGAGATGGAGCAACAGAGTATGCAATTCCTCCCTGTTCA 1346
1141 CTACAAACCTTCCACTACTGAGATGGAGCAACAGAGTATGCAATTCCTCCCTGTTCA 1200
1347 GGAATCTATGAGAGAGAGACTCCGAAACAGTCTCTGATTAAGAGAGAGAGAGATCCCA 1406
1201 GGAATCTATGAGAGAGAGACTCCGAAACAGTCTCTGATTAAGAGAGAGAGAGATCCCA 1260
1407 GAACTGTGAGAGAGAGAGCTTCAAGAGCCAGAGCAAGAGTTCATTCCTGCTGAGAAC 1466
1261 GAACTGTGAGAGAGAGAGCTTCAAGAGCCAGAGCAAGAGTTCATTCCTGCTGAGAAC 1320
1467 CAGACAGATGAGAGAGAGAGAGAGTCCCTTTATATCTTCACTGATCGAAGTGGAGAGCT 1526
1321 CAGACAGATGAGAGAGAGAGAGAGTCCCTTTATATCTTCACTGATCGAAGTGGAGAGCT 1380
1527 GGAGAGCAATTAACACAGAGCTTCTTTTGGCTTTTCCACAGAGAGAGAGAGAGCTTCA 1586
1381 GGAGAGCAATTAACACAGAGCTTCTTTTGGCTTTTCCACAGAGAGAGAGAGAGCTTCA 1440
1587 GAACTGTGAGAGAGAGAGCTTCAAGAGCTTCAAGAGTTCATTCCTGCTGAGAAC 1646
1441 GAACTGTGAGAGAGAGAGCTTCAAGAGCTTCAAGAGTTCATTCCTGCTGAGAAC 1500
1647 TACCTCTTCCCTGAGAGAGAGAGAGTATTTTGGCAAGAGTCTTCACTTCACTTCTGC 1706
1501 TACCTCTTCCCTGAGAGAGAGAGAGTATTTTGGCAAGAGTCTTCACTTCACTTCTGC 1560
1707 CTGAGCATCTATGAGAGAGAGAGTTCAGTCTGCTGAGAGAGAGAGAGAGAGCTTCA 1766
1561 CTGAGCATCTATGAGAGAGAGAGTTCAGTCTGCTGAGAGAGAGAGAGAGAGCTTCA 1620
1767 TTEGAGAGACCAAGCTATTTCTGTGCGAGAGCGCGAGAAAGCAAGTGAAGAGCTGAC 1826
1621 TTEGAGAGACCAAGCTATTTCTGTGCGAGAGCGCGAGAAAGCAAGTGAAGAGCTGAC 1680
1827 GCGGCGAGAGTGGAG 1886
1681 GCGGCGAGAGTGGAG 1740
1887 AATGAGATTTTGGAG 1946

Dh 1741 AATGGAATTGGAGAGCATCATGTCAAGAAACAGGTCAACGGAAAGCTGGGAAAGT 1800
Qy 1947 GGGCAGTCAGTCTAGCTTTTGGGAGCATGAAATCATTTAGGTCTCC 1995
Dh 1801 GGGCAGTCAGTCTAGCTTTTGGGAGCATGAAATCATTTAGGTCTCC 1849
RESULT 19
AX099933 2118 bp DNA linear PAT 02-APR-2001
LOCUS Sequence 15 from Patent WO0120004.
DEFINITION AX099933
ACCESSION AX099933.1 GI:13538943
VERSION
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Yue, H., Tang, Y. T., Bandman, O., Hillman, J. L., Baughn, M. R.,
Azimzai, Y. and Lu, D. A.
TITLE Protein phosphatase and kinase proteins
JOURNAL Patent: WO 0120004-A 15 22-MAR-2001;
Incyte Genomics, Inc. (US)
FEATURES
source 1. .2118
Location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="Incyte ID NO: 1234795CB1"
ORIGIN
Query Match 79.8%; Score 1593; DB 6; Length 2118;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 3; Gaps 2;
Qy 1 ATGGCCCATGAGATGAATGAACTCAATTTGTTACTGAGAGGTGTGCTGTCTGAGAA 60
Dh 25 ATGGCCCATGAGATGAATGAACTCAATTTGTTACTGAGAGGTGTGCTGTCTGAGAA 84
Qy 61 AGTGGAAACGGAAAAAGTGTCTGCTAATTTGATAGCCGGCCATTTGTGGAATACATACATCC 120
Dh 85 AGTGGAAACGGAAAAAGTGTCTGCTAATTTGATAGCCGGCCATTTGTGGAATACATACATCC 144
Qy 121 CACATTTTGGAGCCATTATATATCACTGCTCCAGCTTATGAAAGGAAAGTTGCAACAG 180
Dh 145 CACATTTTGGAGCCATTATATATCACTGCTCCAGCTTATGAAAGGAAAGTTGCAACAG 204
Qy 181 GACAAAGTGTATTAACAGAGCTCATTCAGCATTCAGCGAAACATTAAGTTGACATTTGAT 240
Dh 205 GACAAAGTGTATTAACAGAGCTCATTCAGCATTCAGCGAAACATTAAGTTGACATTTGAT 264
Qy 241 TGCAGTCAGAGGTTGTAGTTTACGATCAAGAGCTCCCAAGATTTGCTCTCTCTTCA 300
Dh 265 TGCAGTCAGAGGTTGTAGTTTACGATCAAGAGCTCCCAAGATTTGCTCTCTCTTCA 324
Qy 301 GACTGTTTCTCACTGTAATCTTGGGTAACTGGAAGAGCTTCAACTCTGTGACCTG 360
Dh 325 GACTGTTTCTCACTGTAATCTTGGGTAACTGGAAGAGCTTCAACTCTGTGACCTG 384
Qy 361 CTTCAGAGTGGGTTTGTCTGAGTTCTCTGTTGTTTCCCTGGCTGTGTAAGAAATCC 420
Dh 385 CTTCAGAGTGGGTTTGTCTGAGTTCTCTGTTGTTTCCCTGGCTGTGTAAGAAATCC 444
Qy 421 ACTCTAGTCCCTTACCTGCACTTTCTAGCCTTCTTACCTGTTGCCAACAATTGGGCCAAC 480
Dh 445 ACTCTAGTCCCTTACCTGCACTTTCTAGCCTTCTTACCTGTTGCCAACAATTGGGCCAAC 504
Qy 481 GCAATTTCTCCGAATCTTATCTTGGCTGCCAGAGAGTGTCTCAACAAGAGCTGTATA 540
Dh 505 GCAATTTCTCCGAATCTTATCTTGGCTGCCAGAGAGTGTCTCAACAAGAGCTGTATG 564
Qy 541 CAGCAGAAATGGAGTTATGTATGTAAATGCAAGCTATACCTGTCCAAAGCCTGACTTT 600

Dh 565 CAGCAGAAATGGAGTTGTTATGTATGTAAATGCAAGAAATACCTGTCCAAAGCCTGACTTT 624
Qy 601 ATCCCCGAGTCTCATTTTCTGTGCTGTGCTGTGAATGACAGCTTTTGTGAAGAAATTTTG 660
Dh 625 ATCCCCGAGTCTCATTTTCTGTGCTGTGCTGTGAATGACAGCTTTTGTGAAGAAATTTTG 684
Qy 661 CCGTGGTTGGAACAAATCAGTAGATTTCATTTGAGAAAGAAAGCCCTCAATGATGTGT 720
Dh 685 CCGTGGTTGGAACAAATCAGTAGATTTCATTTGAGAAAGAAAGCCCTCAATGATGTGT 744
Qy 721 CTAGTCACTGTTTATGCTGGATCTCCGCTCCGCAACATGCTATGCTTACATCATG 780
Dh 745 CTAGTCACTGTTTATGCTGGATCTCCGCTCCGCAACATGCTATGCTTACATCATG 804
Qy 781 AAGAGATGAGCATGTCTTTAGATGAATGACATTCAGATTTTGTGAAGAAAGAAAGCTTACT 840
Dh 805 AAGAGATGAGCATGTGTCTTTAGATGAATGACATTCAGATTTTGTGAAGAAAGAAAGCTTACT 864
Qy 841 ATATCTCCAACTTCAATTTTCTGGGCCAATCTCTGGAATGAGAGAGATTAAAGAC 900
Dh 865 ATATCTCCAACTTCAATTTTCTGGGCCAATCTCTGGAATGAGAGAGATTAAAGAC 924
Qy 901 CAGACTGAGCATCAGGCGCAAGAGCAAACTCAAGCTGTGCACTGTGAGAGGCAAAAT 960
Dh 925 CAGACTGAGCATCAGGCGCAAGAGCAAACTCAAGCTGTGCACTGTGAGAGGCAAAAT 984
Qy 961 GAACCTGTCTCTGCTGTCTCAGAGGGGTGAGAGAGAAACGAGAGACGCGCTTCACTCACCC 1020
Dh 985 GAACCTGTCTCTGCTGTCTCAGAGGGGTGAGAGAGAAACGAGAGACGCGCTTCACTCACCC 1044
Qy 1021 TGTGCCGACTGTGCTACCTCAGAGGCGAGAGAGCAAAAGCCCGTGCATCCGCGCAGCGTG 1080
Dh 1045 TGTGCCGACTGTGCTACCTCAGAGGCGAGAGAGCAAAAGCCCGTGCATCCGCGCAGCGTG 1104
Qy 1081 CCCAGCGTGCCAGCGTGCAGCCGCTGCTGTTAGAGAGCAGCCCGCTGTATCAGGCGCTC 1140
Dh 1105 CCCAGCGTGCCAGCGTGCAGCCGCTGCTGTTAGAGAGCAGCCCGCTGTATCAGGCGCTC 1164
Qy 1141 AGTGGGCTGCACCTGTCCGAGACAGGCTGGAAGACAGCAATTAAGCTCAAGCTTCTTC 1200
Dh 1165 AGTGGGCTGCACCTGTCCGAGACAGGCTGGAAGACAGCAATTAAGCTCAAGCTTCTTC 1224
Qy 1201 TCTCTGATATCAATCAGTTTCAATATATCAGCCAGCATGAGAGATCCTTACATGCGCTTC 1260
Dh 1225 TCTCTGATATCAATCAGTTTCAATATATCAGCCAGCATGAGAGATCCTTACATGCGCTTC 1284
Qy 1261 TCTCTATCAGAAAGATGCTTTGGAATATCAACAACCTTCCACTACTGTGATGAGACCAAC 1320
Dh 1285 TCTCTATCAGAAAGATGCTTTGGAATATCAACAACCTTCCACTACTGTGATGAGACCAAC 1344
Qy 1321 AAGCTATGCCAGTTCTCCCTGTTTCAAGAACTTATGAGAGCACTCCGAAACCAAGT--C 1378
Dh 1345 AAGCTATGCCAGTTCTCCCTGTTTCAAGAACTTATGAGAGCACTCCGAAACCAAGTTC 1404
Qy 1379 CTGATTAAGAGAGAGAGCAGCATCCCAAGAGCTGAGACGCGCAGAGCTTCAAGACGCC 1438
Dh 1405 CTGATTAAGAGAGAGAGCAGCATCCCAAGAGCTGAGACGCGCAGAGCTTCAAGACGCC 1464
Qy 1439 AAGAGAAAGCAATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 1498
Dh 1465 AAGAGAAAGCAATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTCAGTTC 1524
Qy 1499 TATCTCCATCAGATCAGAGTGGAGGCTGTGAGAGCAATTAACAACAAGCTTCTTTTGG 1558
Dh 1525 TATCTCCATCAGATCAGAGTGGAGGCTGTGAGAGCAATTAACAACAAGCTTCTTTTGG 1584
Qy 1559 GCTTTTCCACAGCAGAGAGCACTCAAGAGTGTGCTGAGGCTGAGG--CCTTAAGGCTGG 1617
Dh 1585 GCTTTTCCACAGCAGAGAGCACTCAAGAGTGTGCTGAGGCTGAGGCTTAAAGGCTGG 1644
Qy 1618 CACTCGATATCTTGGCCGCCAAGACTTCACTTCCCTTCCGAGCAGAGCTGTATTTT 1677

```
Db      1645 CACTCGAATATCTTGGCCCCCAGACCTCTACCCCTTCCCTGACCAGAGCTGGTATTTT 1704
Qy      1678 GCCACAGAGTCTCTACACTTCTACTGTGCTCAGCCATCTACGGAGGAGAGTGGCAGTTAC 1737
Db      1705 GCCACAGAGTCTCTACACTTCTACTGTGCTCAGCCATCTACGGAGGAGAGTGGCAGTTAC 1764
Qy      1738 TCTGCCCTACAGCTGCAGCCAGCTGCCCATCTTGGGAGACCAAGTCTATTCTGTGCGCAGG 1797
Db      1765 TCTGCCCTACAGCTGCAGCCAGCTGCCCATCTTGGGAGACCAAGTCTATTCTGTGCGCAGG 1824
Qy      1798 CGGCAAGAGCCAAAGTACAGAGCTGACTCGCGGCGGAGCTGGCATGAGAGAGCCCTTT 1857
Db      1825 CGGCAAGAGCCAAAGTACAGAGCTGACTCGCGGCGGAGCTGGCATGAGAGAGCCCTTT 1884
Qy      1858 GAAAAGCAGTTTAAAGCAGAGAGCTGCCAAATGGAAATTGAGAGAGCATCATGTCAGAG 1917
Db      1885 GAAAAGCAGTTTAAAGCAGAGAGCTGCCAAATGGAAATTGAGAGAGCATCATGTCAGAG 1944
Qy      1918 AACAGGTCACGGAGAGAGCTGGGAAAGTGGGAGTCAAGTCTAGCTTTTCGGGACAGATG 1977
Db      1945 AACAGGTCACGGAGAGAGCTGGGAAAGTGGGAGTCAAGTCTAGCTTTTCGGGACAGATG 2004
Qy      1978 GAAATCATTGAGGTCTCC 1995
Db      2005 GAAATCATTGAGGTCTCC 2022
```

Search completed: June 22, 2004, 14:05:22
Job time : 12411 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

Search completed: June 22, 2004, 16:13:28
Job time : 7677 secs

OM nucleic - nucleic search, using sw model

Run on: June 22, 2004, 09:21:35 ; Search time 7676 Seconds

(without alignments)
7761.210 Million cell updates/sec

Title: US-10-029-345a-108_COPY_538_2532

Perfect score: 1995
Sequence: 1 atggcccatgagatgatgg.....tggaaatcatgaggtctcc 1995

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 1419

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 75 summaries

Database :

EST:*

- 1: em_estda:*
- 2: em_estdm:*
- 3: em_estin:*
- 4: em_estml:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_hic:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_hic:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: em_gss_hum:*
- 18: em_gss_inv:*
- 19: em_gss_pln:*
- 20: em_gss_vrt:*
- 21: em_gss_fun:*
- 22: em_gss_mam:*
- 23: em_gss_mus:*
- 24: em_gss_pro:*
- 25: em_gss_rod:*
- 26: em_gss_phg:*
- 27: em_gss_vrl:*
- 28: gb_gss1:*
- 29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	DB	ID	Description
No. matches found							

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2004, 03:44:37 ; Search time 1140 Seconds

(without alignments)
7434.347 Million cell updates/sec

Title: US-10-029-345A-108_COPY_538_2532

Sequence: 1 atggcccatgagatgattgg.....tggaaatcattgaggtctcc 1995

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3373863 seqs, 212409041 residues

Word size : 1419

Total number of hits satisfying chosen parameters: 28

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 75 summaries

Database : N_Geneseq_29Jan04:*

1: geneseqn1980a:.*
2: geneseqn1990a:.*
3: geneseqn2000a:.*
4: geneseqn2001a:.*
5: geneseqn2001b:.*
6: geneseqn2002a:.*
7: geneseqn2003a:.*
8: geneseqn2003b:.*
9: geneseqn2003c:.*
10: geneseqn2004a:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1995	100.0	5450	6	ACC60559 Polynucle
2	1995	100.0	5450	6	ACC60572 Polynucle
3	1893	94.9	3059	6	AAS15768 Aas15768 cDNA sequ
4	1893	94.9	3496	6	ABK47596 ABK47596 cDNA
5	1893	94.9	3544	5	AAS14639 Aas14639 Human cDN
6	1893	94.9	3544	5	ABK49402 ABK49402 cDNA enco
7	1893	94.9	3766	6	ABK14474 ABK14474 Human gen
8	1893	94.9	4790	6	ABN83966 ABN83966 Human gen
9	1893	94.9	5145	5	ABV20833 ABV20833 Human pro
10	1893	94.9	5145	5	ABV21080 ABV21080 Human pro
11	1893	94.9	5145	5	ABV26680 ABV26680 Human pro
12	1893	94.9	5145	5	ABV20978 ABV20978 Human pro
13	1893	94.9	5145	5	ABV21092 ABV21092 Human pro
14	1893	94.9	5145	5	ABV21312 ABV21312 Human pro
15	1893	94.9	5145	5	ABV21316 ABV21316 Human pro
16	1893	94.9	5145	5	ABV26826 ABV26826 Human pro
17	1893	94.9	5145	5	ABV27131 ABV27131 Human pro
18	1893	94.9	5145	5	ABV26923 ABV26923 Human pro
19	1893	94.9	5145	5	ABV27135 ABV27135 Human pro
20	1893	94.9	5145	5	ABV28657 ABV28657 Human pro
21	1893	94.9	5145	5	ABV22827 ABV22827 Human pro
22	1893	94.9	5145	5	ABV26934 ABV26934 Human pro
23	1849	92.7	2966	4	AAB99685 Aab99685 Human pro

24	1842	92.3	2732	4	AAD09492 AAD09492 Human SGP
25	1842	92.3	3104	6	ABN59704 ABN59704 Novel hum
26	1791	89.8	2102	7	ADA53105 ADA53105 Human cod
27	1791	89.8	5111	6	ACC60521 ACC60521 Polynucle
28	1593	79.8	2118	4	AAF30479 AAF30479 Human pro

ALIGNMENTS

RESULT 1	ACC60559	standard; cDNA; 5450 BP.
ID	ACC60559	
XX	ACC60559;	
AC	ACC60559;	
XX	19-JUN-2003 (first entry)	
XX	Polynucleotide relating to the invention SEQ ID NO: 108.	
XX	Gene; sei; antiproliferative; hepatotropic; nephrotropic; antiarthritic;	
XX	antiproliferative; cardiant; cytosolic; gene therapy; liver disease;	
XX	proliferative disorder; renal failure; cardiovascular disorder;	
XX	immunological disorder; arthritis; psoriasis; congenital heart defect;	
XX	congestive heart failure; cancer; human; mouse; yeast; rat; fruitfly.	
XX	Homo sapiens.	
XX	MO200257460-A2.	
XX	25-JUL-2002.	
XX	20-DEC-2001; 2001WO-US050459.	
XX	20-DEC-2000; 2000US-0256868P.	
XX	30-MAR-2001; 2001US-0280186P.	
XX	01-MAY-2001; 2001US-0287735P.	
XX	05-JUN-2001; 2001US-0295848P.	
XX	25-JUN-2001; 2001US-0300465P.	
XX	(BRIM) BRISTOL-MYERS SQUIBB CO.	
XX	Jackson DG, Feder J, Nelson T, Mintier G, Ramamathan C, Lee L;	
XX	Siemers N, Bol D, Schieven G, Finger J, Todderud CG, Bassolino D;	
XX	Kryetex S, Mcatee P, Suchard S, Banas D;	
XX	WPI; 2002-599721/64.	
XX	P-PsDB; ABR52381.	
PT	Novel polynucleotides encoding human phosphatase polypeptides useful in	
PT	the prevention or treatment of e.g. proliferative and cardiovascular	
PT	disorders.	
XX	Claim 1; Fig 13; 801pp; English.	
CC	The invention relates to a novel isolated nucleic acid comprising a	
CC	polynucleotide having a nucleotide sequence selected from 40	
CC	polynucleotides fully defined in the specification. The polynucleotide of	
CC	the invention has antiproliferative, hepatotropic, nephrotropic,	
CC	antiarthritic, antiproliferative, cardiant, and cytosolic activity. The	
CC	polynucleotide may have a use in gene therapy. A polynucleotide or	
CC	polypeptide of the invention is useful for preventing, treating or	
CC	ameliorating a medical condition, e.g. a proliferative disorder. They are	
CC	also useful for treating e.g. liver disease, renal failure, immunological	
CC	disorders including arthritis and psoriasis, cardiovascular disorders	
CC	such as congenital heart defects and congestive heart failure, and	
CC	cancer. A method of the invention is useful for diagnosing a pathological	
CC	condition or susceptibility to a condition in a subject. The present	
CC	sequence is used in the exemplification of the invention	
XX	Sequence 5450 BP; 1471 A; 1211 C; 1337 G; 1431 T; 0 U; 0 Other;	
XX	Query Match	100.0%; Score 1995; DB 6; Length 5450;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 1995; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 ATGGCCCATGAGATATTGGAATCAATTGTTACTGAGAGTTGGTGGCTCTGCGAA 60
Db 538 ATGGCCCATGAGATATTGGAATCAATTGTTACTGAGAGTTGGTGGCTCTGCGAA 597
OY 61 AGTGAACGGAAGAAAGTCTGCTAATGATAGCCGGCCATTGGTGAATACAAATACATCC 120
Db 598 AGTGAACGGAAGAAAGTCTGCTAATGATAGCCGGCCATTGGTGAATACAAATACATCC 657
OY 121 CACATTTTGAAGCCATTAAATATCAACTGCTCCAACTTATGAAAGCAAGTTTCAACAG 180
Db 658 CACATTTTGAAGCCATTAAATATCAACTGCTCCAACTTATGAAAGCAAGTTTCAACAG 717
OY 181 GACAAAGGTTAATTAACAGAGCTATCCAGATTCAGCGAAACATAAGTTGACATTGAT 240
Db 718 GACAAAGGTTAATTAACAGAGCTATCCAGATTCAGCGAAACATAAGTTGACATTGAT 777
OY 241 TGCAGTCAGAAAGTTAGTTAGTATCAAGTCCCAAGATGTTGCTCTCTCTCA 300
Db 778 TGCAGTCAGAAAGTTAGTTAGTATCAAGTCCCAAGATGTTGCTCTCTCTCTCA 837
OY 301 GACTGTTTCTCACTGTAATCTTGGGTAACTGGAGAAAGCTTCACTCTGTTCACTG 360
Db 838 GACTGTTTCTCACTGTAATCTTGGGTAACTGGAGAAAGCTTCACTCTGTTCACTG 897
OY 361 CTTCAGAGTGGGTTTGGTGAAGTCTCGTGTGTTTCCCTGGCCTCTGTGAAGAAATCC 420
Db 898 CTTCAGAGTGGGTTTGGTGAAGTCTCGTGTGTTTCCCTGGCCTCTGTGAAGAAATCC 957
OY 421 ACTCTAGTCCCTAACCCTGATTTCTACGCTTGTCTTACCTGTTCCCAACATTTGGGCAAC 480
Db 958 ACTCTAGTCCCTAACCCTGATTTCTACGCTTGTCTTACCTGTTCCCAACATTTGGGCAAC 1017
OY 481 CGAATCTTCCCAATCTTTATCTTGGCTGCAAGCAAGATGCTCTCAACAGAGCTGATA 540
Db 1018 CGAATCTTCCCAATCTTTATCTTGGCTGCAAGCAAGATGCTCTCAACAGAGCTGATA 1077
OY 541 CAGCAGATGGAATGTTATGTTAAATGCCAGCTATACCTGTCCAAAGCCTGACTT 600
Db 1078 CAGCAGATGGAATGTTATGTTAAATGCCAGCTATACCTGTCCAAAGCCTGACTT 1137
OY 601 ATCCCGAGTCTCATATTTCTGCGTGTGCTGTGAATGACAGCTTTTGTGAGAAATTTTG 660
Db 1138 ATCCCGAGTCTCATATTTCTGCGTGTGCTGTGAATGACAGCTTTTGTGAGAAATTTTG 1197
OY 661 CCGTGTGGAACAATCAGTAGATTTGATTGAGAAAGCAAAAGCCTCCATGATGTGTT 720
Db 1198 CCGTGTGGAACAATCAGTAGATTTGATTGAGAAAGCAAAAGCCTCCATGATGTGTT 1257
OY 721 CTAGTCACTGTTTAGCTGGATCTCCGCTCCGCAACCATGCTATGCGCTCAATCANG 780
Db 1258 CTAGTCACTGTTTAGCTGGATCTCCGCTCCGCAACCATGCTATGCGCTCAATCANG 1317
OY 781 AAGAGATGGAATGCTTTTAAATGAAGTTTACAGATTTTGAAGAAAGAAAGCCTTACT 840
Db 1318 AAGAGATGGAATGCTTTTAAATGAAGTTTACAGATTTTGAAGAAAGAAAGCCTTACT 1377
OY 841 ATATCTCCAAACTTCAATTTTCTGGGCAACCTCCGAGCTATGAGAAAGAAAGTTTAAAGAC 900
Db 1378 ATATCTCCAAACTTCAATTTTCTGGGCAACCTCCGAGCTATGAGAAAGAAAGTTTAAAGAC 1437
OY 901 CAGACTGAGCATCAGGGCCAAAGCAAACTCAAGTGTGCACTTGGAGAGCCAAAT 960
Db 1438 CAGACTGAGCATCAGGGCCAAAGCAAACTCAAGTGTGCACTTGGAGAGCCAAAT 1497
OY 961 GAACTGTTCCTGCTGTCTCAGAGGTGGAAGAAAGAGAGAGGCTCTCATGTCAACC 1020
Db 1498 GAACTGTTCCTGCTGTCTCAGAGGTGGAAGAAAGAGAGAGGCTCTCATGTCAACC 1557
OY 1021 TGTGCGCACTGTACTCAGAGCAGAGCAAAAGCCGAGTATCCGCGCAAGGTG 1080
```

```
Db 1558 TGTGCCACTCTGCTACTCAAGAGCAGAGACAAGGCCCGGTGATCCGCCAAGCGTG 1617
OY 1081 CCCAGGCTCCCAAGCGTGCAGGCGGTGCTGTTTAAAGGACAGGCCGCTGTTAAGAGCGCTC 1140
Db 1618 CCCAGGCTCCCAAGCGTGCAGGCGGTGCTGTTTAAAGGACAGGCCGCTGTTAAGAGCGCTC 1677
OY 1141 AGTGGGCTGACCTGTCCGCAAGAGGCTGGAAGACAGCAATTAAGCTCAAGGCTTCTTC 1200
Db 1678 AGTGGGCTGACCTGTCCGCAAGAGGCTGGAAGACAGCAATTAAGCTCAAGGCTTCTTC 1737
OY 1201 TCTCTGGAATATCAATACGTTTCAATTCAGGACAGATGAGCATCTTATATGGCTTC 1260
Db 1738 TCTCTGGAATATCAATACGTTTCAATTCAGGACAGATGAGCATCTTATATGGCTTC 1797
OY 1261 TCCATACAGAAAGATGCTTTTGAATACTAACAACTTCCACTCACTCTGATGGAGCAAC 1320
Db 1798 TCCATACAGAAAGATGCTTTTGAATACTAACAACTTCCACTCACTCTGATGGAGCAAC 1857
OY 1321 AAGCTATGCAAGTTCTCCCTGTTTCAAGAACTATCGAGACAGACTCCGAAACCAAGTCT 1380
Db 1858 AAGCTATGCAAGTTCTCCCTGTTTCAAGAACTATCGAGACAGACTCCGAAACCAAGTCT 1917
OY 1381 GATTAAGAGGAAGCAGAGATCCCAAGAAAGCTGAGACCGGCAAGCTTTCAGACAGCCAG 1440
Db 1918 GATTAAGAGGAAGCAGAGATCCCAAGAAAGCTGAGACCGGCAAGCTTTCAGACAGCCAG 1977
OY 1441 AGCAAGCATTTGCAATTCGGTCAGAACCAAGCAGATGCAACCGCCAGAGAGTCCCTTTTA 1500
Db 1978 AGCAAGCATTTGCAATTCGGTCAGAACCAAGCAGATGCAACCGCCAGAGAGTCCCTTTTA 2037
OY 1501 TCTCCATGCAATCCAAAGTGGAGAGGTGAGAGCAATTAACCAACAGACTTCTTTTGGC 1560
Db 2038 TCTCCATGCAATCCAAAGTGGAGAGGTGAGAGCAATTAACCAACAGACTTCTTTTGGC 2097
OY 1561 CTTTCCACCAAGCAGAGCACTTCAAGAAATCTGAGCCGAGGCTTTAAGGAGCTGGCAC 1620
Db 2098 CTTTCCACCAAGCAGAGCACTTCAAGAAATCTGAGCCGAGGCTTTAAGGAGCTGGCAC 2157
OY 1621 TCGATATCTTGGGCCCCCAGACCTTCAACCTTCCCTGACCAAGCAGCTGTAATTTTGGC 1680
Db 2158 TCGATATCTTGGGCCCCCAGACCTTCAACCTTCCCTGACCAAGCAGCTGTAATTTTGGC 2217
OY 1681 ACAGAGTCTCAACTTCTACTGCTGCTCAGCACTTAAGGAGCAGTGCATGTTACTCT 1740
Db 2218 ACAGAGTCTCAACTTCTACTGCTGCTCAGCACTTAAGGAGCAGTGCATGTTACTCT 2277
OY 1741 GCTTACAGCTGCAGCAGCTGCCACTTGGCGAGAACCAAGTCTATTCTGCGCAGAGCGG 1800
Db 2278 GCTTACAGCTGCAGCAGCTGCCACTTGGCGAGAACCAAGTCTATTCTGCGCAGAGCGG 2337
OY 1801 CAGAAAGCAAGTGAACAGAGCTGAATCTGGCGGAGAGCTGSCATGAAGAGAGCCCTTTGAA 1860
Db 2338 CAGAAAGCAAGTGAACAGAGCTGAATCTGGCGGAGAGCTGSCATGAAGAGAGCCCTTTGAA 2397
OY 1861 AAGCAGTTTAAAGCAGAAAGCTGCCAAATGGAATTTGAGAGGCAATGTCAGAGAAC 1920
Db 2398 AAGCAGTTTAAAGCAGAAAGCTGCCAAATGGAATTTGAGAGGCAATGTCAGAGAAC 2457
OY 1921 AAGTCAAGGAGAGAGCTGGGGAAGTGGGCAAGTCAAGTCTTATCTTGGGAGAGCATGGA 1980
Db 2458 AAGTCAAGGAGAGAGCTGGGGAAGTGGGCAAGTCAAGTCTTATGCTTTTGGGAGAGCATGGA 2517
OY 1981 ATCAATTGAGGTCTCC 1995
Db 2518 ATCAATTGAGGTCTCC 2532
```

RESULT 2
ACC60572
ID ACC60572 standard; cDNA; 5450 BP.
XX
ACC60572;
XX

DT 19-JUN-2003 (first entry)
XX Polynucleotide relating to the invention SEQ ID NO: 147.
DE
XX Gene; ss; antiproliferative; hepatotropic; nephrotropic; antiarthritic;
XX antiproliferative; cardiant; cyrostatic; gene therapy; liver disease;
XX proliferative disorder; renal failure; cardiovascular disorder;
XX immunological disorder; arthritis; psoriasis; congenital heart defect;
XX congestive heart failure; cancer; human; mouse; yeast; rat; fruitfly.
XX
OS Homo sapiens.
XX
PN WO200257460-A2.
XX
PD 25-JUL-2002.
XX
PF 20-DEC-2001; 2001WO-US050459.
XX
PR 20-DEC-2000; 2000US-0256668P.
PR 30-MAR-2001; 2001US-0280186P.
PR 01-MAY-2001; 2001US-0287735P.
PR 05-JUN-2001; 2001US-0295848P.
PR 25-JUN-2001; 2001US-0300465P.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
PI Jackson DG, Feder J, Nelson T, Mintier G, Ramanathan C, Lee L,
PI Siemers N, Bol D, Schieven G, Finger J, Todderud CG, Basolino D,
PI Kryetsek S, Mcatee P, Suchard S, Banas D;
XX
DR WPI; 2002-599721/64.
XX
PI P-PSDB; ABR52407.
XX
PT Novel polynucleotides encoding human phosphatase polypeptides useful in
PT the prevention or treatment of e.g. proliferative and cardiovascular
PT disorders.
XX
XX Example 7; Fig 19; 801pp; English.
XX
PS The invention relates to a novel isolated nucleic acid comprising a
XX polynucleotide having a nucleotide sequence selected from 40
CC polynucleotides fully defined in the specification. The polynucleotide of
CC the invention has antiproliferative, hepatotropic, nephrotropic,
CC antiarthritic, antiproliferative, cardiant, and cyrostatic activity. The
CC polynucleotide may have a use in gene therapy. A polynucleotide or
CC polypeptide of the invention is useful for preventing, treating or
CC ameliorating a medical condition, e.g. a proliferative disorder. They are
CC also useful for treating e.g. liver disease, renal failure, immunological
CC disorders including arthritis and psoriasis, cardiovascular disorders
CC such as congenital heart defects and congestive heart failure, and
CC cancer. A method of the invention is useful for diagnosing a pathological
CC condition or susceptibility to a condition in a subject. The present
CC sequence is used in the exemplification of the invention.
XX
SQ Sequence 5450 BP; 1471 A; 1211 C; 1337 G; 1431 T; 0 U; 0 Other;
XX
Query Match 100.0%; Score 1995; DB 6; Length 5450;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 1995; Conservative 0; Mismatches 0; Indels 0; Gaps 0
XX
QY 1 ATGGCCCATGAGATGATTGAACTCAATTTGTACTGAGAGTTGGTGGCTGTCTGGAA 60
DB ATGGCCCATGAGATGATTGAACTCAATTTGTACTGAGAGTTGGTGGCTGTCTGGAA 597
XX
QY 61 AATGGAACGGAAAAAGTGTCTGCTAATTTGATAGCCGGCCATTTTGGAAATTCATCATCC 120
DB AATGGAACGGAAAAAGTGTCTGCTAATTTGATAGCCGGCCATTTTGGAAATTCATCATCC 657
XX
QY 121 CACATTTTGAAGCCATTATATCAATGCTCCAAAGCTTATGAAGGAAGTTGCAACAG 180
DB CACATTTTGAAGCCATTATATCAATGCTCCAAAGCTTATGAAGGAAGTTGCAACAG 717
XX
QY 181 GACCAAGTGTTAATTACAGAGCTCATTCAGCATTCAGCAACATPAAGTTGACATGAT 240

Db	718	GACAAAGTTAATTACGAGCTCATCCGATTCAGGAAACATTAAGCTTGACATGTAT	777
Qy	241	TGCAGTCAGAAAGTTGATGATTAAGATCAAAAGCTCCCAAGATGTGCTCTCTCTTCA	300
Db	778	TGCATCAGAAAGTTGATGATTAAGATCAAAAGCTCCCAAGATGTGCTCTCTCTTCA	837
Qy	301	GACTGTTTCTCACTGATCTTCTGGGTAACCTGAGAAAGACTTCACTGTTCACCTG	360
Db	838	GACTGTTTCTCACTGATCTTCTGGGTAACCTGAGAAAGACTTCACTGTTCACCTG	897
Qy	351	CTTGAGAGTTGGGTTTGCAGATTCTCTCGTTGTTTCCCTGGCCCTGTGGAAGAAATCC	420
Db	898	CTTGAGAGTTGGGTTTGCAGATTCTCTCGTTGTTTCCCTGGCCCTGTGGAAGAAATCC	957
Qy	421	ACTCAGAGCCCACTGATCTTCTCAGGCTTGCTTACCTGTGTGCAATTTGGGCCAAC	480
Db	958	ACTCAGAGCCCACTGATCTTCTCAGGCTTGCTTACCTGTGTGCAATTTGGGCCAAC	1017
Qy	481	CGAATCTTCCCAATCTTTATCTTGCGCTCCAGCGAGATGTCTTCAACAGAGCTGATA	540
Db	1018	CGAATCTTCCCAATCTTTATCTTGCGCTCCAGCGAGATGTCTTCAACAGAGCTGATA	1077
Qy	541	CAGCAGAATGGGATTTGGTATGTGTAAATGCCAGCTTACTCTGCCAAAGCTGACTTT	600
Db	1078	CAGCAGAATGGGATTTGGTATGTGTAAATGCCAGCTTACTCTGCCAAAGCTGACTTT	1137
Qy	601	ATCCCGAGTCATCTTCTCGCGTGCCTGTGATGACAGCTTTTGTGCAAAATTTTG	660
Db	1138	ATCCCGAGTCATCTTCTCGCGTGCCTGTGATGATGACAGCTTTTGTGCAAAATTTTG	1197
Qy	661	CCGTGGTTGACAAATCAGTAGATTTCTTGAAGAAAGCAAAAGCCTCCATGATGTGT	720
Db	1198	CCGTGGTTGACAAATCAGTAGATTTCTTGAAGAAAGCAAAAGCCTCCATGATGTGT	1257
Qy	721	CTAGTGCACTGTTTAGCTGGGATCTCCGCTCCGCACCATCGCTATGCTTACATCATG	780
Db	1258	CTAGTGCACTGTTTAGCTGGGATCTCCGCTCCGCACCATCGCTATGCTTACATCATG	1317
Qy	781	AAGAGGATGACATGCTCTTTAGATGAAGCTTACAGATTTGTGAAGAAAGAAAGCTACT	840
Db	1318	AAGAGGATGACATGCTCTTTAGATGAAGCTTACAGATTTGTGAAGAAAGAAAGCTACT	1377
Qy	841	ATATCTCCAACTTCAATTTTCTGGGCCAATCTCGTGACTATGAGAGAAATTAAGAAC	900
Db	1378	ATATCTCCAACTTCAATTTTCTGGGCCAATCTCGTGACTATGAGAGAAATTAAGAAC	1437
Qy	901	CAGACTGAGCATCAGGGCCAAAGACAACTCAAGCTGTGCACCTTGAGAGGCCAAT	960
Db	1438	CAGACTGAGCATCAGGGCCAAAGACAACTCAAGCTGTGCACCTTGAGAGGCCAAT	1497
Qy	961	GAACTGTCCTCGTGCTCAGAGGGGTGACAGAAAGCGAGAGGCCCTGAGTCAACC	1020
Db	1498	GAACTGTCCTCGTGCTCAGAGGGGTGACAGAAAGCGAGAGGCCCTGAGTCAACC	1557
Qy	1021	TGTGCCGACTGTGCTACTCAGAGGCGACAGACAAAGGCCCTGTGCATCCCGCAGGCTG	1080
Db	1558	TGTGCCGACTGTGCTACTCAGAGGCGACAGACAAAGGCCCTGTGCATCCCGCAGGCTG	1617
Qy	1081	CCAGAGCTGCCAGGCTGACAGCCGTCGTTTGAAGACACGCCGCTGTGTACAGGCTTC	1140
Db	1618	CCAGAGCTGCCAGGCTGACAGCCGTCGTTTGAAGACACGCCGCTGTGTACAGGCTTC	1677
Qy	1141	AGTGGGCTGCACCTGTCCGACAGAGCTGGAAACACAGCAATTAAGCTCAAGCGTTCC	1200
Db	1678	AGTGGGCTGCACCTGTCCGACAGAGCTGGAAACACAGCAATTAAGCTCAAGCGTTCC	1737
Qy	1201	TCTCTGATATCAATCAGTTTCATATTTACAGCCAGCATGTGCAGCATCTTACATGGCTTC	1260
Db	1738	TCTCTGATATCAATCAGTTTCATATTTACAGCCAGCATGTGCAGCATCTTACATGGCTTC	1797
Qy	1261	TCTCTATCAGAAAGTCTTTGGAATCTACAAACCTTCCACTACTTGTGATGGACCAAC	1320

Db 1798 TCCTCATCAGAAAGTCTTGGAACTACTACAACTTCCACTACTGTGATGGAGCAAC 1857
QY 1321 AAGCTATGCCAGTCTCTCCCTGTTGAGAACTATGAGAGCAAGTCTCCCGAAACCACTCT 1380
Db 1858 AAGCTATGCCAGTCTCTCCCTGTTGAGAACTATGAGAGCAAGTCTCCCGAAACCACTCT 1917
QY 1381 GATTAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1440
Db 1918 GATTAAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1977
QY 1441 AGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1500
Db 1978 AGCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2037
QY 1501 TCTCCATGATGCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560
Db 2038 TCTCCATGATGCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2097
QY 1561 CTTTCCACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1620
Db 2098 CTTTCCACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2157
QY 1621 TCGGATATCTTGGGCCCCCGGAGAGCTTCACTCCCTGTCGACGAGCTGTGATTTTGGC 1680
Db 2158 TCGGATATCTTGGGCCCCCGGAGAGCTTCACTCCCTGTCGACGAGCTGTGATTTTGGC 2217
QY 1681 ACAGAGTCTCTACACTTCTACTGCTGCTGAGCTTCACTGAGGAGGAGGAGGAGGAGGAGGAG 1740
Db 2218 ACAGAGTCTCTACACTTCTACTGCTGCTGAGCTTCACTGAGGAGGAGGAGGAGGAGGAGGAG 2277
QY 1741 GCTTACAGCTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1800
Db 2278 GCTTACAGCTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2337
QY 1801 CAGAGGCAAGTACAGAGTGTACTGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1860
Db 2338 CAGAGGCAAGTACAGAGTGTACTGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2397
QY 1861 AAGCAGTTTAAACGAGAAAGCTGCAATGAAATTTGAGAGAGCATCATGTCAGAGAAC 1920
Db 2398 AAGCAGTTTAAACGAGAAAGCTGCAATGAAATTTGAGAGAGCATCATGTCAGAGAAC 2457
QY 1921 AGGTCAACGGAAGAGTGTGGGAAAGTGGGCAAGTCAAGTCTTTTGGGAGCAATGAA 1980
Db 2458 AGGTCAACGGAAGAGTGTGGGAAAGTGGGCAAGTCAAGTCTTTTGGGAGCAATGAA 2517
QY 1981 ATCATGAGGTCTCC 1995
Db 2518 ATCATGAGGTCTCC 2532

RESULT 3
AAS15768
ID AAS15768 standard; cDNA; 3059 BP.
AC AAS15768;
XX
DT 18-JUN-2002 (first entry)
XX
DE cDNA sequence of human (dual specificity phosphatase) DUSP-10.
XX
KW Human; dual specificity phosphatase; DUSP-10; cancer; epilepsy; stroke;
neurological degeneration syndrome; Alzheimer's disease; depression;
schizophrenia; asthma; immune disorder; inflammatory process; arthritis;
osteoporosis; diabetes; gene therapy; chromosome 12; ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 127..2124
FT /tag= a
FT /product= "DUSP-10 protein"

PN W0200177340-A1.
XX 18-OCT-2001.
PD 06-APR-2001; 2001WO-EP003966.
XX
PF 10-APR-2000; 2000EP-00107143.
XX
PR (MERCK) MERCK PATENT GMBH.
XX
PA Duecker K;
XX
PI WPI; 2002-010917/01.
XX P-PSTB; A4U09946.
DR
PT Novel dual specificity phosphatase polypeptides useful for treating
PT cancers, epilepsy, stroke, neuronal degeneration syndromes, Alzheimer's
PT disease, depression, schizophrenia, asthma and immune disorders.
PS Claim 5; Page 34-37; 43pp; English.

CC The present invention relates to a new isolated dual specificity
CC phosphatase (DUSP10) polypeptide, comprising a 665 residue amino acid
CC sequence that is fully defined in the specification. The invention also
CC provides a sequence encoded by a 3059 nucleotide sequence fully defined
CC in the specification, and a sequence having at least 95 % identity to the
CC polypeptide, or fragments or variants of DUSP-10. The invention is useful
CC for treating cancer e.g. leukemia, colon carcinoma, lung cancer,
CC prostate cancer, metastasis of tumour cells, neo-angiogenesis, epilepsy,
CC stroke, neuronal degeneration syndromes, Alzheimer's disease, depression,
CC schizophrenia, cardiac myopathies, asthma, immune disorders,
CC inflammatory processes e.g. arthritis, bowel disease, type I diabetes,
CC osteoporosis, diabetes and diabetes associated diseases. The molecules of
CC the invention are also useful as vaccines for inducing immunological
CC response in a mammal, in disease diagnosis and in assays for screening
CC agonistic or antagonistic compounds. Other uses of the invention include
CC identifying membrane bound or soluble receptors, as a diagnostic reagent,
CC in chromosome localization studies, and as a valuable tool in tissue
CC expression studies. The present sequence represents cDNA of the human
CC dual specificity phosphatase, DUSP-10. DUSP-10 is located on chromosome
CC 12
XX
SQ Sequence 3059 BP; 831 A; 761 C; 709 G; 758 T; 0 U; 0 Other;

Query Match 94.9%; Score 1893; DB 6; Length 3059;
Best local similarity 99.9%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCCCATGAGATGATTTGGAACCTCAATGTTTACTGAGAGGTTGGTGGCTGGTGA 60
Db 127 ATGGCCCATGAGATGATTTGGAACCTCAATGTTTACTGAGAGGTTGGTGGCTGGTGA 186
QY 61 AGTGAACGGAAGAGTGTGCTGCTAATTTGATGACCGGCAATTTTGGAAATACATCC 120
Db 187 AGTGAACGGAAGAGTGTGCTGCTAATTTGATGACCGGCAATTTTGGAAATACATCC 246
QY 121 CACATTTTGAAGGCATTAATATCACTGCTCCAGAGTTTGAAGCAAGTGTGCAACAG 180
Db 247 CACATTTTGAAGGCATTAATATCACTGCTCCAGAGTTTGAAGCAAGTGTGCAACAG 306
QY 181 GACAAAGTGTAAATTACAGAGCTATCCAGCATTCAGCGAAACATAAGTTGACATTGAT 240
Db 307 GACAAAGTGTAAATTACAGAGCTATCCAGCATTCAGCGAAACATAAGTTGACATTGAT 366
QY 241 TGCAGTCAGAAAGTTGATTTACATCAAGTATCCCAAGATGTGCTCTCTCTTCA 300
Db 367 TGCAGTCAGAAAGTTGATTTACATCAAGTATCCCAAGATGTGCTCTCTCTTCA 426
QY 301 GACTGTTTCTCACTGACTCTGGGTAACATGGAAGAGCTTCAACTCTGTTCACCTG 360
Db 427 GACTGTTTCTCACTGACTCTGGGTAACATGGAAGAGCTTCAACTCTGTTCACCTG 486
QY 361 CTTCAGAGTGGTTTGTGATGATCTCTGTTTCTCTGGCTCTGTGAAGGAATCC 420

Db 487 CTGGAGGAGGGTGTGCTGAGTCTCTCGTTGTTTCCCTGGCTCTGTGAAGAAATATC 546
 Qy 421 ACTCTAGTCCCTACCTGCAATTTTTCAGCTCTGTTACTCTGTTCCCAATTTGGCCAAAC 480
 Db 547 ACTCTAGTCCCTACCTGCAATTTTTCAGCTCTGTTACTCTGTTCCCAATTTGGCCAAAC 606
 Qy 481 CGAATCTTCCCAATTTTATCTTGGCTGCAAGGAGATGCTCTCAACAGAGCTGATA 540
 Db 607 CGAATCTTCCCAATTTTATCTTGGCTGCAAGGAGATGCTCTCAACAGAGCTGATA 666
 Qy 541 CAGCAAAATGGGATTTGTTATGTTAAATGCAAGCTATACCTGTCCAAAGCTGACTTT 600
 Db 667 CAGCAAAATGGGATTTGTTATGTTAAATGCAAGCTATACCTGTCCAAAGCTGACTTT 726
 Qy 601 ATCCCGAGTCTGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
 Db 727 ATCCCGAGTCTGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 786
 Qy 661 CCGTGTGGAGCAAAATCAGTAAATTTTCAATTTGAGAAAGCAAAAGCTCCCAATGATGTT 720
 Db 787 CCGTGTGGAGCAAAATCAGTAAATTTTCAATTTGAGAAAGCAAAAGCTCCCAATGATGTT 846
 Qy 721 CTAGTCACTGTTTAACTGAGATCTCCGCTCCGCAACCATGCTATGCTCATCATG 780
 Db 847 CTAGTCACTGTTTAACTGAGATCTCCGCTCCGCAACCATGCTATGCTCATCATG 906
 Qy 781 AAGAGATGACATGTTCTTAAATGAGTTCAGATTTTGTGAAGAAAAAGAAAGCACTTACT 840
 Db 907 AAGAGATGACATGTTCTTAAATGAGTTCAGATTTTGTGAAGAAAAAGAAAGCACTTACT 966
 Qy 841 ATATCTCCCAATTTTAACTGAGATCTCCGCTCCGCAACCATGCTATGCTCATCATG 900
 Db 967 ATATCTCCCAATTTTAACTGAGATCTCCGCTCCGCAACCATGCTATGCTCATCATG 1026
 Qy 901 CAGACTGAGCATCAGGAGCAAAAGCAAACTCAAGCTGCTGCACTGAGAGAGCAAAAT 960
 Db 1027 CAGACTGAGCATCAGGAGCAAAAGCAAACTCAAGCTGCTGCACTGAGAGAGCAAAAT 1086
 Qy 961 GAACTGTCTCTGCTCTCTCAAGAGGTGAGCAAGAAAGCAAGCCCTCACTGATCAACCC 1020
 Db 1087 GAACTGTCTCTGCTCTCTCAAGAGGTGAGCAAGAAAGCAAGCCCTCACTGATCAACCC 1146
 Qy 1021 TGTGCGGATCTCTGCTCTCTCAAGAGGTGAGCAAGAAAGCAAGCCCTCACTGATCAACCC 1080
 Db 1147 TGTGCGGATCTCTGCTCTCTCAAGAGGTGAGCAAGAAAGCAAGCCCTCACTGATCAACCC 1206
 Qy 1081 CCGAGGAGTCCGAGGAGTCCGCTGCTGTTAGAGGAGCAAGCCGCTGAGTCAAGAGGCTTC 1140
 Db 1207 CCGAGGAGTCCGAGGAGTCCGCTGCTGTTAGAGGAGCAAGCCGCTGAGTCAAGAGGCTTC 1266
 Qy 1141 AGTGGGCTGCACTGTCTCCGAGACAGGCTGGAAGACAGCAATTAAGTCAAGCGTCTCTTC 1200
 Db 1267 AGTGGGCTGCACTGTCTCCGAGACAGGCTGGAAGACAGCAATTAAGTCAAGCGTCTCTTC 1326
 Qy 1201 TCTCTGAGATTCAAATCAAGTTTATATTTAGAGGAGTGGAGAGATCTTCACTGCTTC 1260
 Db 1327 TCTCTGAGATTCAAATCAAGTTTATATTTAGAGGAGTGGAGAGATCTTCACTGCTTC 1386
 Qy 1261 TCTCTCATGAGAAATGTTGAAATCTCAAAACCTTCCACTCTGATGAGGAGCAAC 1320
 Db 1387 TCTCTCATGAGAAATGTTGAAATCTCAAAACCTTCCACTCTGATGAGGAGCAAC 1446
 Qy 1321 AAGCTATGCAAGTTCTCCCTGTTTCAAGAACTATGAGAGCACTCCGAAACCACTGCT 1380
 Db 1447 AAGCTATGCAAGTTCTCCCTGTTTCAAGAACTATGAGAGCACTCCGAAACCACTGCT 1506
 Qy 1381 GATTAAGAGAGAGCAGCATCTCCCAAGAACTGCAAGCCGAGGCTTCAAGACACGAG 1440
 Db 1507 GATTAAGAGAGAGCAGCATCTCCCAAGAACTGCAAGCCGAGGCTTCAAGACACGAG 1566
 Qy 1441 AGCAAGCATTTGATTCGATTCAGAAACAGAGAGTGGAGCCGAGAGAGTCCCTTTTA 1500

Db 1567 AGCAAGCATTCATTCGCTCAGAAACAGACAGTGGACCGCCAGAGATCCCTTTTA 1626
 Qy 1501 TCTCCATGCAATCAGAAATGGAGCGGTGAGCAATTTACACACAGCTTCTTTTGGC 1560
 Db 1627 TCTCCATGCAATCAGAAATGGAGCGGTGAGCAATTTACACACAGCTTCTTTTGGC 1686
 Qy 1561 CTTTCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1620
 Db 1687 CTTTCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1746
 Qy 1621 TCGATATCTGAGTCCGCTGAGTCCCTTCCCTGACAGCAGCAGTGTATTTTGGC 1680
 Db 1747 TCGATATCTGAGTCCGCTGAGTCCCTTCCCTGACAGCAGCAGTGTATTTTGGC 1806
 Qy 1681 ACAGAGTCTCACTTCTACTCTGCTGAGCAGTATAGAGAGCAGTGTATTTTGGC 1740
 Db 1807 ACAGAGTCTCACTTCTACTCTGCTGAGCAGTATAGAGAGCAGTGTATTTTGGC 1866
 Qy 1741 GCTACAGCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1800
 Db 1867 GCTACAGCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1926
 Qy 1801 CAGAAAGCAAGTACAGAGCTGATCTGCGCGAGAGCTGAGTGAAGAGAGCCCTTTGAA 1860
 Db 1927 CAGAAAGCAAGTACAGAGCTGATCTGCGCGAGAGAGCTGAGTGAAGAGAGCCCTTTGAA 1986
 Qy 1861 AAGCAATTTAAACGCAAGAGCTGCCAATTTGAGAGAGCTATGTCAGAGAAC 1920
 Db 1987 AAGCAATTTAAACGCAAGAGCTGCCAATTTGAGAGAGCTATGTCAGAGAAC 2046
 Qy 1921 AGGTACGAGGAGAGAGCTGGGAGAAAGTGGAGAGTCTAGCTTTTGGGAGCATGAA 1980
 Db 2047 AGGTACGAGGAGAGAGCTGGGAGAAAGTGGAGAGTCTAGCTTTTGGGAGCATGAA 2106
 Qy 1981 ATCATTGAGTCTCC 1995
 Db 2107 ATCATTGAGTCTCC 2121

RESULT 4
 ABR47596
 ID ABR47596 standard; cDNA; 3496 BP.
 XX
 AC ABR47596;
 XX
 DT 02-JUL-2002 (first entry)
 XX
 DE cDNA encoding human dual-specificity phosphatase-3 (DSP-16) protein.
 XX
 KW Human; dual-specificity phosphatase-3; DSP-16; autoimmune disease;
 KW mitogen-activated protein kinase; MAP; Duchenne muscular dystrophy;
 KW cancer; graft-versus-host disease; allergy; metabolic disease;
 KW abnormal cell growth; abnormal cell proliferation; contact inhibition;
 KW cell cycle abnormality; anchorage independent cell growth; apoptosis;
 KW intercellular adhesion; DSP-16 modulator; chromosome 12p; gene; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 562..2559
 FT /tag=a
 FT /product="human dual-specificity phosphatase-3 (DSP-16)
 FT protein"
 FT
 PD WO200226997-A2.
 XX
 PD 04-APR-2002.
 XX
 PF 25-SEP-2001; 2001WO-US030124.
 XX
 PR 26-SEP-2000; 2000US-0235487P.
 XX
 PA (CEPT-) CEPTYR INC.

XX Luche RM, Wei B;
 XX WPI: 2002-315802/35.
 DR P-PSDB: MAU79156.
 XX
 XX New DSP-16 polypeptide, useful for identifying modulators of its
 FT activity, which can be used in the treatment of disorders such as
 PT Duchenne muscular dystrophy, or cancer.
 XX
 PS Claim 7; Fig 1; 87bp; English.
 CC The present invention relates to a new polypeptide, DSP-16, having a 665
 CC amino acid sequence, given in the specification, or a variant having at
 CC least 50 % identical residues, which retains the ability to
 CC dephosphorylate an activated mitogen-activated protein (MAP) kinase. The
 CC invention can be used for identifying agents which modulate DSP-16
 CC activity; for modulation of a proliferative response in a cell, survival
 CC of a cell, or differentiation of a cell. The cell displays contact
 CC inhibition of cell growth or anchorage independent growth and may display
 CC altered intercellular adhesion. The agent may modulate apoptosis, or the
 CC cell cycle. The identified modulators can be used to treat Duchenne
 CC muscular dystrophy, cancer, graft-versus-host disease, autoimmune
 CC diseases, allergies, metabolic diseases, abnormal cell growth, abnormal
 CC cell proliferation, and cell cycle abnormalities. The present nucleic
 CC acid sequence is that of the human dual-specificity phosphatase-3 (DSP-
 CC 16) gene located on chromosome 12p. This sequence encodes the human DSP-
 CC 16 protein of the invention
 XX
 SQ Sequence 3496 BP; 939 A; 854 C; 856 G; 847 T; 0 U; 0 Other;
 Query Match 94.9%; Score 1893; DB 6; Length 3496;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGAGCCATGAGATGATGGAAGTCAATGTTTACTGAGAGGTGTGCTCTGCTGAA 60
 DB 562 ATGAGCCATGAGATGATGGAAGTCAATGTTTACTGAGAGGTGTGCTCTGCTGAA 621
 QY 61 AGTGAACGGAAAAAGTGTGCTAATTGATAGCCGSCATTTGTGAAATACATCATCC 120
 DB 622 AGTGAACGGAAAAAGTGTGCTAATTGATAGCCGSCATTTGTGAAATACATCATCC 681
 QY 121 CACATTTTGAAGCCCTTAATATCAAGTCTCTCAAGCTTATGAAGGAGGTGGACAG 180
 DB 682 CACATTTTGAAGCCCTTAATATCAAGTCTCTCAAGCTTATGAAGGAGGTGGACAG 741
 QY 181 GACAAAGTGTAAATACAGAGCTCATCCAGCATTCAGCGAAACATTAAGTTGACTTAT 240
 DB 742 GACAAAGTGTAAATACAGAGCTCATCCAGCATTCAGCGAAACATTAAGTTGACTTAT 801
 QY 241 TGCAGTCAGAGGTTGTAGTTTACATCAAGTCCCAAGATGTGCTCTCTCTTCA 300
 DB 802 TGCAGTCAGAGGTTGTAGTTTACATCAAGTCCCAAGATGTGCTCTCTCTTCA 861
 QY 301 GACTGTTTCTACCTGTACTTCTGGGTAACTGAGAGAGCTTCAACTCTGTTCACCTG 360
 DB 862 GACTGTTTCTACCTGTACTTCTGGGTAACTGAGAGAGCTTCAACTCTGTTCACCTG 921
 QY 361 CTGAGAGGTGTGTGCTGAGTTCTCTGTTGTTCCCTGGCTCTGTGAGAGAAATCC 420
 DB 922 CTGAGAGGTGTGTGCTGAGTTCTCTGTTGTTCCCTGGCTCTGTGAGAGAAATCC 981
 QY 421 ACTTAGTCCCTACCTGCAATTTCTGAGCTTGCTTACCTGTGCAACATTTGGGCAAC 480
 DB 982 ACTTAGTCCCTACCTGCAATTTCTGAGCTTGCTTACCTGTGCAACATTTGGGCAAC 1041
 QY 481 CGAATTTCTCCCAATCTTATCTTGGGCTGCCAGAGATGCTCTCAACAGAGCTGATA 540
 DB 1042 CGAATTTCTCCCAATCTTATCTTGGGCTGCCAGAGATGCTCTCAACAGAGCTGATA 1101
 QY 541 CAGAGAAATGGGATTTATGTATGTTAAATGCGAGCTTACCTGTCGAAGCTGACCTT 600

DB 1102 CAGAGAAATGGGATTTATGTATGTTAAATGCGAGCAATACCTGTCCAAAGCTGACTTT 1161
 QY 601 ATCCCGAGTCTCATTTCTGTGCTGCTGTGAATGACAGCTTTTGTGAGAAATTTTG 660
 DB 1162 ATCCCGAGTCTCATTTCTGTGCTGCTGTGAATGACAGCTTTTGTGAGAAATTTTG 1221
 QY 661 CCGTGTGAGAAATTCAGTATGATTTCAATGAGAAAGCAAAAGCTCCATGATGTGTT 720
 DB 1222 CCGTGTGAGAAATTCAGTATGATTTCAATGAGAAAGCAAAAGCTCCATGATGTGTT 1281
 QY 721 CTAGTGCACTGTTTGTAGTGGATCTCCGCTCCGCAACATGCTATGCTCAATCANG 780
 DB 1282 CTAGTGCACTGTTTGTAGTGGATCTCCGCTCCGCAACATGCTATGCTCAATCANG 1341
 QY 781 AAGAGATGACATGCTTTTATGATGATGATCAAGATTTGTGAAAAGAAAAAGCTTACT 840
 DB 1342 AAGAGATGACATGCTTTTATGATGATGATCAAGATTTGTGAAAAGAAAAAGCTTACT 1401
 QY 841 ATATCTCCAAATCTCAATTTTCTGGGCCAATCTCTGCACTATGAGAGAAATTAAGAAC 900
 DB 1402 ATATCTCCAAATCTCAATTTTCTGGGCCAATCTCTGCACTATGAGAGAAATTAAGAAC 1461
 QY 901 CAGACTGAGATCAGGCGCAAGCAAACTCAAGCTGTGCACTGAGAGAGCAAT 960
 DB 1462 CAGACTGAGATCAGGCGCAAGCAAACTCAAGCTGTGCACTGAGAGAGCAAT 1521
 QY 961 GAACTGTCTCTGCTGTCTTCAAGAGGTGAGACAAAAGAGAGAGCTCTCACTCAACC 1020
 DB 1522 GAACTGTCTCTGCTGTCTTCAAGAGGTGAGACAAAAGAGAGAGCTCTCACTCAACC 1581
 QY 1021 TGTGCGCATCTGTGCACTCTCAGAGCAGAGCAAAAGGCGCTGTGATCCCGCAGAGGTG 1080
 DB 1582 TGTGCGCATCTGTGCACTCTCAGAGCAGAGCAAAAGGCGCTGTGATCCCGCAGAGGTG 1641
 QY 1081 CCAAGCTGTCCAGCTGTGAGAGCGCTGTGATGAGAGAGAGCTGTGATGAGAGAGCTG 1140
 DB 1642 CCAAGCTGTCCAGCTGTGAGAGCGCTGTGATGAGAGAGAGCTGTGATGAGAGAGCTG 1701
 QY 1141 AGTGGGCTGCACTGTGCTGCGAGAGAGAGCTGTGAGAGAGAGCTGTGAGAGAGCTG 1200
 DB 1702 AGTGGGCTGCACTGTGCTGCGAGAGAGAGAGCTGTGAGAGAGAGCTGTGAGAGAGCTG 1761
 QY 1201 TCTCTGATATCAAAATCAATGATTTCAATTCAGCCAGATGAGAGATCTTACATGCTTC 1260
 DB 1762 TCTCTGATATCAAAATCAATGATTTCAATTCAGCCAGATGAGAGATCTTACATGCTTC 1821
 QY 1261 TCTCTATCAGAAATGCTTTTGAATTAATCAAACTTCACTGATGAGAGAGCAAC 1320
 DB 1822 TCTCTATCAGAAATGCTTTTGAATTAATCAAACTTCACTGATGAGAGAGCAAC 1881
 QY 1321 AAGCTATGCAAGTCTCCCTGTTCAGGAATTAATGAGAGAGCTCCGAAACCAATGCTT 1380
 DB 1882 AAGCTATGCAAGTCTCCCTGTTCAGGAATTAATGAGAGAGCTCCGAAACCAATGCTT 1941
 QY 1381 GATTAAGAGAGAGCAAGATCCCAAGAGCTCAGACGCGAGGCTTTCAGAGAGCAAG 1440
 DB 1942 GATTAAGAGAGAGCAAGATCCCAAGAGCTCAGACGCGAGGCTTTCAGAGAGCAAG 2001
 QY 1441 AGCAAGGATGATGATTTGGTCAAGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGT 1500
 DB 2002 AGCAAGGATGATGATTTGGTCAAGACCAAGAGAGAGAGAGAGAGAGAGAGAGT 2061
 QY 1501 TCTCACTGATGAGAGTGGAGAGCTGTGAGAGCAATTAACACAGAGCTTCTTTTCCGGC 1560
 DB 2062 TCTCACTGATGAGAGTGGAGAGCTGTGAGAGCAATTAACACAGAGCTTCTTTTCCGGC 2121
 QY 1561 CTTTCCACAGCAGCAGAGACCTCAAGAGCTGTGAGAGCTGTGAGAGAGAGAGAGAG 1620
 DB 2122 CTTTCCACAGCAGCAGAGACCTCAAGAGCTGTGAGAGCTGTGAGAGAGAGAGAGAG 2181
 QY 1621 TCGGATATCTTGGAG 1680
 DB 2182 TCGGATATCTTGGAG 2241

QY 1681 ACAGAGTCCACACTTCTACTGCTGACCATCTAGAGGACAGTGCAGTACTCTCT 1740
DB 2242 ACAGAGTCCACACTTCTACTGCTGACCATCTAGAGGACAGTGCAGTACTCTCT 2301
QY 1741 GCCTACAGCTGACAGCAGTGTCCCACTTGTGCGAGACCAAGTCTATTCTGTGCGCAGGCG 1800
DB 2302 GCCTACAGCTGACAGCAGTGTCCCACTTGTGCGAGACCAAGTCTATTCTGTGCGCAGGCG 2361
QY 1801 CAGAGCCAAAGTACAGAGTGTACTGCGCGGAGCTGCGATTAAGAGAGCCCTTTGAA 1860
DB 2362 CAGAGCCAAAGTACAGAGTGTACTGCGCGGAGCTGCGATTAAGAGAGCCCTTTGAA 2421
QY 1861 AAGCAGTTTAAAGCGAAGCTGCCAAATGGAATTTGAGAGAGCATCATGTGAGAGAC 1920
DB 2422 AAGCAGTTTAAAGCGAAGCTGCCAAATGGAATTTGAGAGAGCATCATGTGAGAGAC 2481
QY 1921 AGGTACAGGGAAGAGTGGGGAAGTGGGAGTCAAGTCTAGTCTTTTCGGGACAGATGAA 1980
DB 2482 AGGTACAGGGAAGAGTGGGGAAGTGGGAGTCAAGTCTAGTCTTTTCGGGACAGATGAA 2541
QY 1981 ATCATTGAGGTCTCC 1995
DB 2542 ATCATTGAGGTCTCC 2556

RESULT 5
AAS14639
ID AAS14639 standard; cDNA; 3544 BP.
XX
AC AAS14639;
XX
DT 18-DEC-2001 (first entry)
DE Human cDNA encoding dual specificity phosphatase 21117.
XX
KW Human; as: dual specificity phosphatase 21117; hepatotropic; cytosolic;
KW hemato-oncologic disorder; autoimmune disorder; diabetes mellitus;
KW rheumatoid arthritis; multiple sclerosis; Crohn's disease;
KW liver disorder; erythroid associated disorder; haemolytic anaemia;
KW cellular proliferative; differential disorder; leukaemia;
KW acute myeloid leukaemia; carcinoma; sarcoma; metastatic cancer.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 589..2586
FT /tag= a
FT /product= "Dual specificity phosphatase 21117"
XX
XX MO200173059-A2.
XX
XX 04-OCT-2001.
XX
XX 23-MAR-2001; 2001MO-US009477.
XX
XX 24-MAR-2000; 2000US-0191858P.
XX
XX (MILL-) MILLENNIUM PHARM INC.
XX
XX Meyers RA;
XX
XX WPI, 2001-611635/70.
XX
XX P-PSDB; AAU09016.
XX
XX New human dual specificity polypeptides and nucleic acids for diagnosis
XX PT of disease and treatment of e.g. liver disorders.
XX
XX Claim 1; Fig 1; 143pp; English.
XX
XX The invention relates to two novel human dual specificity phosphatases
XX CC designated 21117 and 38692, the nucleic acids encoding them (including
XX CC fragments, allelic variants, their complements or nucleic acids that

CC hybridise to them) and antibodies raised against the proteins. The
CC antibody is useful for detecting the presence of the polypeptide, and the
CC nucleic acid fragments are useful for detecting the presence of the
CC nucleic acid, for diagnosing disease. Agents (e.g. the antibody and
CC antisense sequences) are useful for modulating the activity or expression
CC of the polypeptide/nucleic acid, useful for the treatment of e.g.
CC hematopoietic disorders, autoimmune disorders (e.g. diabetes mellitus,
CC rheumatoid arthritis, multiple sclerosis, Crohn's disease and many others
CC listed in the specification) liver disorders, erythroid associated
CC disorders (e.g. haemolytic anaemia) cellular proliferative or
CC differentiative disorders, leukaemias (e.g. acute myeloid leukaemia),
CC carcinomas, sarcomas, metastatic cancers. Agents modulating 21117 and
CC 38692 are also useful for modulating the proliferation, survival,
CC migration or differentiation of a 38692 or 21117-expressing cell. The
CC polypeptide and nucleic acids are useful for identifying modulating
CC agents. The present sequence encodes the dual specificity phosphatase
CC 21117
XX
SQ Sequence 3544 BP; 1000 A; 840 C; 822 G; 882 T; 0 U; 0 Other;
XX
Query Match 94.9%; Score 1893; DB 5; Length 3544;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGGCCCATGAGATGATGATGGAATCAATTTGATGAGAGTGGTGGCTGTGAA 60
DB 589 ATGGCCCATGAGATGATGATGGAATCAATTTGATGAGAGTGGTGGCTGTGAA 648
QY 61 AGTGAACGGAAAAAGTGTGCTAATTTGATGAGAGTGGTGGCTGTGAA 120
DB 649 AGTGAACGGAAAAAGTGTGCTAATTTGATGAGAGTGGTGGCTGTGAA 708
QY 121 CACATTTTGAAGCCATTAATCACTGCTCCAACTTATGAGGAAAGTTGCAAC 180
DB 709 CACATTTTGAAGCCATTAATCACTGCTCCAACTTATGAGGAAAGTTGCAAC 768
QY 181 GACAAAGTGTATTAACAGAGCTCATCCAGCATTCAGCGAACAATAGGTTGATGAT 240
DB 769 GACAAAGTGTATTAACAGAGCTCATCCAGCATTCAGCGAACAATAGGTTGATGAT 828
QY 241 TGCAGTCAGAAAGTGTGATTTACATCAAGTCTCCCAAGATGTTGCTCTCTTCA 300
DB 829 TGCAGTCAGAAAGTGTGATTTACATCAAGTCTCCCAAGATGTTGCTCTCTTCA 888
QY 301 GACTGTTTTCTCACTGTAATCTTGGGTAATGGAAGAGCTTCAACTGTTCACCTG 360
DB 889 GACTGTTTTCTCACTGTAATCTTGGGTAATGGAAGAGCTTCAACTGTTCACCTG 948
QY 361 CTTCAGAGTGGGTTGCTGAGTCTCTGTTGTTTCCCTGGGCTGTGGAAGAAATCC 420
DB 949 CTTCAGAGTGGGTTGCTGAGTCTCTGTTGTTTCCCTGGGCTGTGGAAGAAATCC 1008
QY 421 ACTTAGTCCCTACCTGCAATTTCTCAGCCTTGCTTACTGTTGCAAACTTGGGCCAAC 480
DB 1009 ACTTAGTCCCTACCTGCAATTTCTCAGCCTTGCTTACTGTTGCAAACTTGGGCCAAC 1068
QY 481 CGAATTTCTCCCAATTTTATCTTGGCTGCAAGAGATGCTCTCAACAGAGCTGATG 540
DB 1069 CGAATTTCTCCCAATTTTATCTTGGCTGCAAGAGATGCTCTCTCAACAGAGCTGATG 1128
QY 541 CAGCAGATGGAGATGGTATGTTAATGAGCAGCTATACCTGTCACCAAGCCTGACTTT 600
DB 1129 CAGCAGATGGAGATGGTATGTTAATGAGCAGCTATACCTGTCACCAAGCCTGACTTT 1188
QY 601 ATCCCGAGTCTCATTTCTGCGTGTGCTGCTGTAATGACAGCTTTTGTGAGAAATTTTG 660
DB 1189 ATCCCGAGTCTCATTTCTGCGTGTGCTGCTGTAATGACAGCTTTTGTGAGAAATTTTG 1248
QY 661 CCGTGGTGAACAATCAGTATGATTTGATTTGATGAGAAAGCAAAAGCTTCAATGATGTT 720
DB 1249 CCGTGGTGAACAATCAGTATGATTTGATTTGATGAGAAAGCAAAAGCTTCAATGATGTT 1308
QY 721 CTAAGTCACTGTTTAGCTGGAGATCTCCGCTCCGCAACATGCTATCGCTACATCATG 780

```
Db 1309 CTAGGCACTGTTAGCTGGGATCTCCGCTCCGCAACATCGCTATCGCTCAATCATG 1368
Qy 781 AAGAGATGACATCTCTTTAGATGAGACTTACAGATTTTGAAAAGAAAAGACTTACT 840
Db 1369 AAGAGATGACATCTCTTTAGATGAGACTTACAGATTTTGAAAAGAAAAGACTTACT 1428
Qy 841 ATATCTCCAACTTCAATTTTCTGGGCAACTCTCGGACTATGAGAAAGATTAAGAAC 900
Db 1429 ATATCTCCAACTTCAATTTTCTGGGCAACTCTCGGACTATGAGAAAGATTAAGAAC 1488
Qy 901 CAGACTGAGCATCAGGGCCAAAGAGCAAACTCAAGCTGTGCACTTGAGAAAGCAAAAT 960
Db 1489 CAGACTGAGCATCAGGGCCAAAGAGCAAACTCAAGCTGTGCACTTGAGAAAGCAAAAT 1548
Qy 961 GAACTGTCCCTGCTGTCTCAGAGGGGTGACAGAAAAGCCAGAGCCCTCACTCCACCC 1020
Db 1549 GAACTGTCCCTGCTGTCTCAGAGGGGTGACAGAAAAGCCAGAGCCCTCACTCCACCC 1608
Qy 1021 TGTGCCGACTCTGTCTACCTCAGAGGCAAGCAAAAGCCCTGTGCACTCCGCGCAGGCTG 1080
Db 1609 TGTGCCGACTCTGTCTACCTCAGAGGCAAGCAAAAGCCCTGTGCACTCCGCGCAGGCTG 1668
Qy 1081 CCCAGCGTCCGCGCAGGCTGTGCAAGCCGTCTGTGAGAGCAAGCCGCTGTGACAGGCGCTC 1140
Db 1669 CCCAGCGTCCGCGCAGGCTGTGCAAGCCGTCTGTGAGAGCAAGCCGCTGTGACAGGCGCTC 1728
Qy 1141 AGTGGGCTGACCTGTCTCCGAGACAGGCTGGAAGACAGCAATAAGCTCAAGGCTTCTTTC 1200
Db 1729 AGTGGGCTGACCTGTCTCCGAGACAGGCTGGAAGACAGCAATAAGCTCAAGGCTTCTTTC 1788
Qy 1201 TCTCTGAGATTCAAATCAGTTTCATATTCAGCCAGATGAGCATCTTACATGAGCTTC 1260
Db 1789 TCTCTGAGATTCAAATCAGTTTCATATTCAGCCAGATGAGCATCTTACATGAGCTTC 1848
Qy 1261 TCTCTGAGATTCAAATCAGTTTCATATTCAGCCAGATGAGCATCTTACATGAGCTTC 1320
Db 1849 TCTCTGAGATTCAAATCAGTTTCATATTCAGCCAGATGAGCATCTTACATGAGCTTC 1908
Qy 1321 AAGCTATGCGCAGTTCTCCCTGTTTCAAGAACTATCGGAGACAGCTCCGAAACCAAGTCTT 1380
Db 1909 AAGCTATGCGCAGTTCTCCCTGTTTCAAGAACTATCGGAGACAGCTCCGAAACCAAGTCTT 1968
Qy 1381 GATTAAGAGAGAACCGACATCCCAAGAAAGCTGACAGCCGAGGCTTACAGACAGCAG 1440
Db 1969 GATTAAGAGAGAACCGACATCCCAAGAAAGCTGACAGCCGAGGCTTACAGACAGCAG 2028
Qy 1441 AGCAAGCAGATGATTCGCTGAGAACCAAGCAGAGGCAAGCCGAGGCTTACAGACAGCAG 1500
Db 2029 AGCAAGCAGATGATTCGCTGAGAACCAAGCAGAGGCAAGCCGAGGCTTACAGACAGCAG 2088
Qy 1501 TCTTCACTGATCGAAGTGGAGCGGTGAGCAATTAACAACAGCAGCTTCTTTTGGGC 1560
Db 2089 TCTTCACTGATCGAAGTGGAGCGGTGAGCAATTAACAACAGCAGCTTCTTTTGGGC 2148
Qy 1561 CTTTCCACAGCAGCAGACACTTCAAGAAAGTCTGCTGGCTTGAAGGCTTGAAGGCTGAGC 1620
Db 2149 CTTTCCACAGCAGCAGACACTTCAAGAAAGTCTGCTGGCTTGAAGGCTTGAAGGCTGAGC 2208
Qy 1621 TCGGATATCTTGGCCCCCAGACCTTACCCCTTCCGACACAGAGATGATTTTGGCC 1680
Db 2209 TCGGATATCTTGGCCCCCAGACCTTACCCCTTCCGACACAGAGATGATTTTGGCC 2268
Qy 1681 ACAGAGTCTTCACTTCTTACTCTGCTCAGCCATCTACGAGGAGCAGCAGTTACTCT 1740
Db 2269 ACAGAGTCTTCACTTCTTACTCTGCTCAGCCATCTACGAGGAGCAGCAGTTACTCT 2328
Qy 1741 GCCTACAGTCAAGCAGCAGCTGCCACTTGGCGAGACCAAGTCTATTTCTGTGCGAGCGG 1800
Db 2329 GCCTACAGTCAAGCAGCAGCTGCCACTTGGCGAGACCAAGTCTATTTCTGTGCGAGCGG 2388
Qy 1801 CAGAAGCCAGTGAAGAGCTGACTCGGCGGAGCTGGCATGAAGAGAGCCCTTTGAA 1860
```

```
Db 2389 CAGAACCCAGTGCAGACGCTACTCGCGCGAGCTGGCATGAAGAGCCCTTTGAA 2448
Qy 1861 AAGCACTTTAAACGAGAACTGCCAAATGGAATTTGAGAGAGCATGTCAGAGAAC 1920
Db 2449 AAGCACTTTAAACGAGAACTGCCAAATGGAATTTGAGAGAGCATGTCAGAGAAC 2508
Qy 1921 AGGTCAAGGAGAGAGCTGGGAAAGTGGGCACTGCTTACTTTTGGGAGCATGGA 1980
Db 2509 AGGTCAAGGAGAGAGCTGGGAAAGTGGGCACTGCTTACTTTTGGGAGCATGGA 2568
Qy 1981 ATCATTGAGGTCTCC 1995
Db 2569 ATCATTGAGGTCTCC 2583

RESULT 6
ABK49402
ID ABK49402 standard; cDNA; 3544 BP.
XX
XX ABK49402;
AC
XX
XX 02-JUL-2002 (first entry)
DT
XX
XX cDNA encoding human dual specificity phosphatase 21117 protein.
DE
XX
XX Human; dual specificity phosphatase 21117; erythroid-related disorder;
KW haematopoietic-related disorder; leukaemia; autoimmune disease; anaemia;
KW erythrocytosis; liver-related disorder; cancer; gene; ss.
XX
XX Homo sapiens.
OS
XX
XX
XX Key Location/Qualifiers
FH 589..2586
FT CDS /*tag= a
FT /product= "Human dual specificity phosphatase 21117"
FT /note= "Specifically claimed in claim 2"
PN
XX
XX US2002034807-A1.
PD
XX 21-MAR-2002.
XX
XX 23-MAR-2001; 2001US-00816494.
PF
XX
XX 24-MAR-2000; 2000US-0191858P.
PR
XX
XX (MEYE/) MEYERS R A.
PA
XX
XX Meyers RA;
PI
XX
XX WPI; 2002-351088/38.
DR
XX
XX P-PSDB; AAU79929.
DT
XX
XX New nucleic acids, designated 36692 and 21117, encoding dual specificity
PT phosphatases for treating cell proliferation and differentiation
PT disorders including hematopoietic and erythroid-related disorders and
PS cancers.
XX
XX Claim 2; Fig 1; 76pp; English.
XX
XX The present invention relates to new nucleic acids designated 36692 and
XX 21117 encoding dual specificity phosphatase family members. The nucleic
XX acid, polypeptide encoded by it, and antibody specific for the
XX polypeptide may be used to diagnose and treat hematopoietic-related
XX disorders such as leukaemias and autoimmune diseases, erythroid-related
XX disorders such as anaemias and erythrocytosis, liver-related disorders,
XX and cancers, particularly of the breast, colon, adipose, prostate and
XX lung. The present nucleic acid sequence encodes the human dual
XX specificity phosphatase 21117 protein of the invention, as described
XX above
SQ
Sequence 3544 BP; 1000 A; 840 C; 822 G; 882 T; 0 U; 0 Other;
Query Match 94.9%; Score 1893; DB 6; Length 3544;
```


Best Local Similarity 99.9%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
QY 1 ATGGCCCATGAGATGATGGAATCAATATTGTAAGAGAGTTGGTGGCTCTGCTGAA 60
DB 589 ATGGCCCATGAGATGATGGAATCAATATTGTAAGAGTTGGTGGCTCTGCTGAA 648
QY 61 AGTGAACGGAAGAGTGTCTAATTGATAGCCGGCAATTTGGAAATACATATAC 120
DB 649 AGTGAACGGAAGAGTGTCTAATTGATAGCCGGCAATTTGGAAATACATATAC 708
QY 121 CACATTTTGGAGCCATTAATATCACTGCTCCAGCTTAATGAAGGAGTTGCAAC 180
DB 709 CACATTTTGGAGCCATTAATATCACTGCTCCAGCTTAATGAAGGAGTTGCAAC 768
QY 181 GACAAAGTGTAAATTCAGAGCTCATCCAGCATTCAGCGAAATTAAGTTGAT 240
DB 769 GACAAAGTGTAAATTCAGAGCTCATCCAGCATTCAGCGAAATTAAGTTGAT 828
QY 241 TGGAGTCAGAGGTTGATTAAGATCAAGATCCCAAGATGTTGGCTCTCTCTGA 300
DB 829 TGGAGTCAGAGGTTGATTAAGATCAAGATCCCAAGATGTTGGCTCTCTCTGA 888
QY 301 GACTGTTTCTCACTGACTTCTGGGTAACGTGAGAGAGCTTCACTCTGTTCA 360
DB 889 GACTGTTTCTCACTGACTTCTGGGTAACGTGAGAGAGCTTCACTCTGTTCA 948
QY 361 CTTCAGAGTGGGTTTGTGAGTTCTCTGTTGTTCCCTGGCTCTGTGAGAGAAATCC 420
DB 949 CTTCAGAGTGGGTTTGTGAGTTCTCTGTTGTTCCCTGGCTCTGTGAGAGAAATCC 1008
QY 421 ACTCTAGTCCCTACCTGCAATTTCTCAGCCCTTGCTTACCTGTGCCAATTTGGCCCAAC 480
DB 1009 ACTCTAGTCCCTACCTGCAATTTCTCAGCCCTTGCTTACCTGTGCCAATTTGGCCCAAC 1068
QY 481 CGAATTCCTCCCACTTTTATCTTGGCTGCACAGAGATGCTCTCAACAAGAGCTGATA 540
DB 1069 CGAATTCCTCCCACTTTTATCTTGGCTGCACAGAGATGCTCTCAACAAGAGCTGATA 1128
QY 541 CAGCAGATGGAGTTGGTTATGTTTAATGCAAGCTATACCTGTCCAAAGCCTGACTT 600
DB 1129 CAGCAGATGGAGTTGGTTATGTTTAATGCAAGCTATACCTGTCCAAAGCCTGACTT 1188
QY 601 ATCCCGAGTCTAATTCCTGCGTGGCTGTAATGACAGCTTTTGTGAAATTTTG 660
DB 1189 ATCCCGAGTCTAATTCCTGCGTGGCTGTAATGACAGCTTTTGTGAAATTTTG 1248
QY 661 CCGTGGTGGCAAAATCAGATATTCATTGAGAAAGCAAAAGCCTCCAAATGATGTT 720
DB 1249 CCGTGGTGGCAAAATCAGATATTCATTGAGAAAGCAAAAGCCTCCAAATGATGTT 1308
QY 721 CTAGTGCATGTTTAACTGGGATCTCCGCTCCGCCAATGCTATGCTCTACATCAT 780
DB 1309 CTAGTGCATGTTTAACTGGGATCTCCGCTCCGCCAATGCTATGCTCTACATCAT 1368
QY 781 AAGAGATGACATGCTTTAGATGAGCTTACAGATTTGTGAAAGAAAAAAGACTACT 840
DB 1369 AAGAGATGACATGCTTTAGATGAGCTTACAGATTTGTGAAAGAAAAAAGACTACT 1428
QY 841 ATATCTCCAAACTTCAATTTTCTGGGCCAATCCTGGAATATGAGAAAGATTTAAGAC 900
DB 1429 ATATCTCCAAACTTCAATTTTCTGGGCCAATCCTGGAATATGAGAAAGATTTAAGAC 1488
QY 901 CAGACTGAGCATCAGGGCCAAAGAGCAACTCAAGCTGTGCACTTGAGAAAGCCAAAT 960
DB 1489 CAGACTGAGCATCAGGGCCAAAGAGCAACTCAAGCTGTGCACTTGAGAAAGCCAAAT 1548
QY 961 GAACTGTCCCTGCTGTCTCAGAGGGTGAACAGAAAAAGCAGAGCCCTCACTCCACCC 1020
DB 1549 GAACTGTCCCTGCTGTCTCAGAGGGTGAACAGAAAAAGCAGAGCCCTCACTCCACCC 1608
QY 1021 TGTGCCGACTCTGTACTCTCAGAGGAGCAGAGCAAAAGGCCGTGCACTCCGCGAGGCTG 1080
```

```
DB 1609 TGTGCCGACTCTGTACTCTCAGAGGAGCAGAGCAAAAGGCCGTGCACTCCGCGAGGCTG 1668
QY 1081 CCCAGGTCGCCAGGCTGAGCGTGTGTTAGAGACAGGCCGCTGGTACAGGGCTC 1140
DB 1669 CCCAGGTCGCCAGGCTGAGCGTGTGTTAGAGACAGGCCGCTGGTACAGGGCTC 1728
QY 1141 AGTGGCTCAGCTGTCCGACAGAGGCTGGAAGACAGCAATTAAGCTCAAGCTTCTTC 1200
DB 1729 AGTGGCTCAGCTGTCCGACAGAGGCTGGAAGACAGCAATTAAGCTCAAGCTTCTTC 1788
QY 1201 TCTCTGATATCAATCACTTATATTCAGCCAGATGCGAGCATCTCTTACATGCTTC 1260
DB 1789 TCTCTGATATCAATCACTTATATTCAGCCAGATGCGAGCATCTCTTACATGCTTC 1848
QY 1261 TCTCTCAAGAAAGCTTTGGAATATCAAACTTCCATCTCTGATGGGAGCAAC 1320
DB 1849 TCTCTCAAGAAAGCTTTGGAATATCAAACTTCCATCTCTGATGGGAGCAAC 1908
QY 1321 AAGCTATGCAAGTCTCCCTGTTCAAGAACTATCGAGACAGATCCCGAAACAGTCT 1380
DB 1909 AAGCTATGCAAGTCTCCCTGTTCAAGAACTATCGAGACAGATCCCGAAACAGTCT 1968
QY 1381 GATTAAGAGAAAGCCAGCATCCCAAGAGCTGACAGCCGCAAGCTTTCAACAGCCAG 1440
DB 1969 GATTAAGAGAAAGCCAGCATCCCAAGAGCTGACAGCCGCAAGCTTTCAACAGCCAG 2028
QY 1441 AGCAAGCATTTGCTTGGTCAAGAACAGACAGATGGGACCGGCCAGAGTCCCTTTA 1500
DB 2029 AGCAAGCATTTGCTTGGTCAAGAACAGACAGATGGGACCGGCCAGAGTCCCTTTA 2088
QY 1501 TCTCCACATCAGAAAGTGGAGCGTGGAGGCAATTAACACAGCAGTCTCTTTCCG 1560
DB 2089 TCTCCACATCAGAAAGTGGAGCGTGGAGGCAATTAACACAGCAGTCTCTTTCCG 2148
QY 1561 CTTCACAGCCAGCAGCAGCCTCAAGAGTCTGCTGAGCTGAGGCTTAAGGGCTGGCAC 1620
DB 2149 CTTCACAGCCAGCAGCAGCCTCAAGAGTCTGCTGAGGCTTAAGGGCTGGCAC 2208
QY 1621 TGGGATATCTTGGCCCCCAGACCTTACCCCTTCCCTCAACAGCAGCTGTATTTTGGC 1680
DB 2209 TGGGATATCTTGGCCCCCAGACCTTACCCCTTCCCTCAACAGCAGCTGTATTTTGGC 2268
QY 1681 ACAGAGTCTCACTCACTTACTCTGCTCAGCCATCTACGAGGAGCTGCACTTACT 1740
DB 2269 ACAGAGTCTCACTCACTTACTCTGCTCAGCCATCTACGAGGAGCTGCACTTACT 2328
QY 1741 GCCTACAGCTGACAGCAGCTGCCACTTGCAGAGCAAGCTTATCTGTGCGCAGGCG 1800
DB 2329 GCCTACAGCTGACAGCAGCTGCCACTTGCAGAGCAAGCTTATCTGTGCGCAGGCG 2388
QY 1801 CAGAGCCAAAGTGAACAGACTGACTGCGCGCGAGCTGGCATGAGAGCCCTTTGAA 1860
DB 2389 CAGAGCCAAAGTGAACAGACTGACTGCGCGCGAGCTGGCATGAGAGCCCTTTGAA 2448
QY 1861 AAGCAGTTTAAACGAGAAAGCTGCCAAATGGAATTTGAGAGAGCATGTCAGAGAAC 1920
DB 2449 AAGCAGTTTAAACGAGAAAGCTGCCAAATGGAATTTGAGAGAGCATGTCAGAGAAC 2508
QY 1921 AGGTCAAGGAAAGAGCTGGGAAAGTGGGCAAGTCAAGTCTTATGCTTTCGCGCAGATGAA 1980
DB 2509 AGGTCAAGGAAAGAGCTGGGAAAGTGGGCAAGTCAAGTCTTATGCTTTCGCGCAGATGAA 2568
QY 1981 ATCATTTAGAGTCTCC 1995
DB 2569 ATCATTTAGAGTCTCC 2583
```

RESULT 7
ABK14474
ID ABK14474 standard; cDNA; 3766 BP.
XX
AC ABK14474;
XX

Db	1678	AGTGGGCTGCACCTGTCCGACAGACGGCTGGAAAGACAACTAAGCTCAAGCTTCTTC	1737
Qy	1201	TCTCTGATATCAAAATCAGTTTCATATTCACGCCAGCATGGAGCATCTTCATATGCTTC	1260
Db	1738	TCTCTGATATCAAAATCAGTTTCATATTCACGCCAGCATGGAGCATCTTCATATGCTTC	1797
Qy	1261	TCCTCATCAGAAAGATGCTTTGGAAATCTACAAACCTTCATCTCTGGATGGGACCAAC	1320
Db	1798	TCCTCATCAGAAAGATGCTTTGGAAATCTACAAACCTTCATCTCTGGATGGGACCAAC	1857
Qy	1321	AAGCTATGCGAGTTCTCCCTGTTTAGGAACTATCGAGCGAGACTCCGAAACCAAGTCT	1380
Db	1858	AAGCTATGCGAGTTCTCCCTGTTTAGGAACTATCGAGCGAGACTCCGAAACCAAGTCT	1917
Qy	1361	GATTAAGAGGAAGCCAGCATCCCAAGAAAGTGCAGACCCGACAGCTTCAGACAGCCAG	1440
Db	1918	GATTAAGAGGAAGCCAGCATCCCAAGAAAGTGCAGACCCGACAGCTTCAGACAGCCAG	1977
Qy	1441	AGCAAGCATTTGCATTCGCTGACGAACCCAGCAGCTGCAACCCGCCAGAGTCCCTTTTA	1500
Db	1978	AGCAAGCATTTGCATTCGCTGACGAACCCAGCAGCTGCAACCCGCCAGAGTCCCTTTTA	2037
Qy	1501	TCTCCACTGCATCGAAAGTGGGAGGTGGAGGACATTTACCAACACGTTCTCTTTTGCC	1566
Db	2038	TCTCCACTGCATCGAAAGTGGGAGGTGGAGGACATTTACCAACACGTTCTCTTTTGCC	2097
Qy	1561	CTTTCCACAGCCAGCAGCAGCACCTCAAGAACTGTGCTGGCTTGGGCTTTAAGGCTGGCAC	1620
Db	2098	CTTTCCACAGCCAGCAGCAGCACCTCAAGAACTGTGCTGGCTTGGGCTTTAAGGCTGGCAC	2157
Qy	1621	TCGGATATCTTTGGCCCCCCCAGACTCTACCCCTTCCTGACACGACTGGTATTTTGGC	1680
Db	2158	TCGGATATCTTTGGCCCCCCCAGACTCTACCCCTTCCTGACACGACTGGTATTTTGGC	2217
Qy	1681	ACAGAGTCTCTACACTTCTACTCTGCGCTCAGCATCTACAGAGAGAGCCAGTTACTCT	1740
Db	2218	ACAGAGTCTCTACACTTCTACTCTGCGCTCAGCATCTACAGAGAGAGCCAGTTACTCT	2277
Qy	1741	GCCTACAGCTGCAGCAGCTGCGCCACTTGGCGGAGCAAGTCTATTCTGTGCGAGCGG	1800
Db	2278	GCCTACAGCTGCAGCAGCTGCGCCACTTGGCGGAGCAAGTCTATTCTGTGCGAGCGG	2337
Qy	1801	CAGAAGCCAAGTACAGAGCTGACTCGCGCGGAGCTGGCATGAAGAGAGCCCTTTGAA	1860
Db	2338	CAGAAGCCAAGTACAGAGCTGACTCGCGCGGAGCTGGCATGAAGAGAGCCCTTTGAA	2397
Qy	1861	AAGCAGTTTAAACGCAAGAGCTGCGCAATTTGGAATTTGGAGAGCATATGTCAGAGAAC	1920
Db	2398	AAGCAGTTTAAACGCAAGAGCTGCGCAATTTGGAATTTGGAGAGCATATGTCAGAGAAC	2457
Qy	1921	AGGTCAACGGGAAGAGCTGGGGGAAATGGGACAGTCACTAGCTTTTGGGGAGCATGGA	1980
Db	2458	AGGTCAACGGGAAGAGCTGGGGGAAATGGGACAGTCACTAGCTTTTGGGGAGCATGGA	2517
Qy	1981	ATCATTTGAGGTCTCC 1995	
Db	2518	ATCATTTGAGGTCTCC 2532	
RESULT 8			
ABN83966			
ID	ABN83966	standard; DNA; 4790 BP.	
XX	ABN83966;		
AC			
XX	06-SEP-2002 (first entry)		
DT			
XX			
DE	Human gene sequence #13.		
XX			
KW	Human; brain; consil; hippocampus; foetal brain; diagnosis; gene; ds.		
OS	Homo sapiens.		

XX	Key	Location/Qualifiers
FH	FT CDS	184..2181 /*tag= a
PN	MO200252005-A1.	
PD	04-JUL-2002.	
PF	20-DEC-2001; 2001WO-JP011217.	
PR	22-DEC-2000; 2000JP-00389742.	
PA	(KAZU-) KAZUSA DNA RES INST FOUND. (CELE-) CELESTAR LEXICO-SCI LTD.	
PI	Ohara O, Nagase T, Nakajima D;	
DR	WPI; 2002-500762/53.	
DR	P-PDB; ABB97946.	
PT	Genes and their expression products cloned from human cDNA libraries for treatment and diagnosis of diseases associated with their expression.	
PS	Claim 1(a); Page 111-117; 238pp; Japanese.	
CC	The invention relates to DNA encoding polypeptides directly cloned from cDNA libraries originating in adult whole brain, human consil, human adult hippocampus and human foetal whole brain. Polypeptides and polynucleotides of the invention may be used in the investigation of differential expression of the DNA sequences in normal subjects and disease patients. They may also be used in the production of antibodies, oligonucleotide probes and DNA chips for diagnosis and identification of drugs for treatment of diseases with which the DNA sequences are associated. The sequences given in records ABN83954-ABN83984 represent human gene sequences of the invention	
SQ	Sequence 4790 BP; 1293 A; 1064 C; 1152 G; 1281 T; 0 U; 0 Other;	
	Query Match 94.9%; Score 1893; DB 6; Length 4790; Best Local Similarity 99.9%; Pred. No. 0; Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0	
OY	1 ATGGCCCATGAGATGATTGAACCTCAATTGTACTGAAGGTTGGTCTGCTGAA 60	
Db	184 ATGGCCCATGAGATGATTGAACCTCAATTGTACTGAAGGTTGGTCTGCTGAA 243	
OY	61 AGTGAACGGAAAAAGTCGTCTAATTGATAGCCGCCATTGTGTGAATACAAATACATCC 120	
Db	244 AGTGAAACGGAAAAAGTCGTCTAATTGATAGCCGCCATTGTGTGAATACAAATACATCC 303	
OY	121 CACAATTTGGAAGCATTATATCAATCGCTCAGAAGCTTATGAGGGAAGTTGCACACAG 180	
Db	304 CACAATTTGGAAGCATTATATCAATCGCTCAGAAGCTTATGAGGGAAGTTGCACACAG 363	
OY	181 GACAAAGTGTAAATTACAGAGCTCACGACATTACGAGAAAATTAAGGTTGACATTGAT 240	
Db	364 GACAAAGTGTAAATTACAGAGCTCACGACATTACGAGAAAATTAAGGTTGACATTGAT 423	
OY	241 TGCACTCAGAAAGTTGATGTTTAGCATCAAAGCTCCCAAGATGTTGCTCTCTCTTCA 300	
Db	424 TGCACTCAGAAAGTTGATGTTTAGCATCAAAGCTCCCAAGATGTTGCTCTCTCTTCA 483	
OY	301 GACGTTTTTCACACTGACTCTTGCGGTAACTGAGAAAGAGCTTCAACTCTGTCACCTG 360	
Db	484 GACGTTTTTCACACTGACTCTTGCGGTAACTGAGAAAGAGCTTCAACTCTCTGTCACCTG 543	
OY	361 CTTCGAGGTGGGTTGCTGAGTTCTCTGTTGTTTCCCTGGCCTCTGTGAAGAAAATCC 420	
Db	544 CTTCGAGGTGGGTTGCTGAGTTCTCTGTTGTTTCCCTGGCCTCTGTGAAGAAAATCC 603	
OY	421 ACTCTAGTCCCTAACCTGCATTTTCAGCCTTGCTTAAGTGTGCAACTTTGGGCCAAC 480	

Db 604 ACTAGTCCCTACCTGATCTTCAGCGCTTGCTTACCTGTTGGCAACATTTGGGCCAAC 663
Qy 481 CGAATCTTCCCAATCTTATCTTGGCTGCGCAGAGGAGTGTCTCAACAGAGAGCTGATA 540
Db 664 CGAATCTTCCCAATCTTATCTTGGCTGCGCAGAGGAGTGTCTCAACAGAGAGCTGATA 723
Qy 541 CAGCAGAAATGGGATTTGGTATGTGTAAATGCACTATACCTGTCCAAAGCCGTGATTT 600
Db 724 CAGCAGAAATGGGATTTGGTATGTGTAAATGCACTATACCTGTCCAAAGCCGTGATTT 783
Qy 601 ATCCCGAGTCTCATTTCTGCGGTGTGCTGTGAATGAACAGCTTTTGTGAATAATTTTG 660
Db 784 ATCCCGAGTCTCATTTCTGCGGTGTGCTGTGAATGAACAGCTTTTGTGAATAATTTTG 843
Qy 661 CCGTGTGTGACAAATCAGTATGATTTGATGAGAAACCAAAAGCTTCATATGATGTT 720
Db 844 CCGTGTGTGACAAATCAGTATGATTTGATGAGAAACCAAAAGCTTCATATGATGTT 903
Qy 721 CTAGTGCATGTTTATGCTGGGATCTCCGCTCCGCCACATCGCTATCGCTACATCATG 780
Db 904 CTAGTGCATGTTTATGCTGGGATCTCCGCTCCGCCACATCGCTATCGCTACATCATG 963
Qy 781 AAGAGATGACATGTCTTTAGATGAAGCTTACAGATTTGTGAAGAAAAGACCTTACT 840
Db 964 AAGAGATGACATGTCTTTAGATGAAGCTTACAGATTTGTGAAGAAAAGACCTTACT 1023
Qy 841 ATATCTCCAACTTCAATTTTCTGGGCCAATCTCGGACCTATGAGAAGAAATTAAGAC 900
Db 1024 ATATCTCCAACTTCAATTTTCTGGGCCAATCTCGGACCTATGAGAAGAAATTAAGAC 1083
Qy 901 CAGACTGGAGCATCAGGGCCAAAGAGCAAACTCAAGCTGTGACCTGAGAGCCCAAT 960
Db 1084 CAGACTGGAGCATCAGGGCCAAAGAGCAAACTCAAGCTGTGACCTGAGAGCCCAAT 1143
Qy 961 GAACTGTGCTGCTGTCTCAGAGGGTGAACAGAAAAGGAGAGCGCCCTCAGTCCACCC 1020
Db 1144 GAACTGTGCTGCTGTCTCAGAGGGTGAACAGAAAAGGAGAGCGCCCTCAGTCCACCC 1203
Qy 1021 TGTGCGCACTGTCTACTCTCAGAGGACAGACAAAGGCCGTGATCCCGCAGCGTG 1080
Db 1204 TGTGCGCACTGTCTACTCTCAGAGGACAGACAAAGGCCGTGATCCCGCAGCGTG 1263
Qy 1081 CCCAGGCTGCCAGCGCTGACGCCGTGTGTAGAGAGACGCCCGCTGTACAGCGCTC 1140
Db 1264 CCCAGGCTGCCAGCGCTGACGCCGTGTGTAGAGAGACGCCCGCTGTGTACAGCGCTC 1323
Qy 1141 AGTGGCTGCACTGTCCGAGACAGGCTGGAAGACAGAAATAGTCAAGGTTCTTC 1200
Db 1324 AGTGGCTGCACTGTCCGAGACAGGCTGGAAGACAGAAATAGTCAAGGTTCTTC 1383
Qy 1201 TCTCTGATATCAATCAGTTTCAATTCAGCCAGCATGGCAGCATCTTACATGCTTC 1260
Db 1384 TCTCTGATATCAATCAGTTTCAATTCAGCCAGATGGCAGCATCTTACATGCTTC 1443
Qy 1261 TCTCTATCGAAGATGCTTTGGAATCTTCAAACTTCCATCTCTGTGATGGACAAC 1320
Db 1444 TCTCTATCGAAGATGCTTTGGAATCTTCAAACTTCCATCTCTGTGATGGACAAC 1503
Qy 1321 AAGCTATGCGAGTTCTCCCTGTTCAGGAATATGAGAGACATCCCGAAACCATGCTCT 1380
Db 1504 AAGCTATGCGAGTTCTCCCTGTTCAGGAATATGAGAGACATCCCGAAACCATGCTCT 1563
Qy 1381 GATTAAGAGAAACCAAGCATCCCAAGAAAGCTGCAAGCCGACGCTTTCAGACAGCAG 1440
Db 1564 GATTAAGAGAAACCAAGCATCCCAAGAAAGCTGCAAGCCGACGCTTTCAGACAGCAG 1623
Qy 1441 AGCAGAGCATGTCATTCGTGAGAACAGCAGCAGCATGGGACCGCCAGAGGTCCTTTTA 1500
Db 1624 AGCAGAGCATGTCATTCGTGAGAACAGCAGCAGCATGGGACCGCCAGAGGTCCTTTTA 1683
Qy 1501 TCTCCATGTCATCGAAGTGGAGCGTGGAGGACATTAACACACAGCTTCTTTTGGGC 1560
Db 1684 TCTCCATGTCATCGAAGTGGAGCGTGGAGGACATTAACACACAGCTTCTTTTGGGC 1743

Qy 1561 CTTTCCACAGCCAGACAGCCTCAAGAGTCTGCTGGCGCTGAGCTTAAAGGCTGGAC 1620
Db 1744 CTTTCCACAGCCAGACAGCCTCAAGAGTCTGCTGGCGCTGAGCTTAAAGGCTGGAC 1803
Qy 1621 TGGGATATCTTGGCCCCCAGACCTCTACCCCTTCTGACCAAGAGCTGTGATTTGCC 1680
Db 1804 TGGGATATCTTGGCCCCCAGACCTCTACCCCTTCTGACCAAGAGCTGTGATTTGCC 1863
Qy 1681 ACAGAGTCTCTACATTTCTACTCTGCTCAGCAGCATCTACGAGGAGGAGCCAGTACTCT 1740
Db 1864 ACAGAGTCTCTACATTTCTACTCTGCTCAGCAGCATCTACGAGGAGGAGCCAGTACTCT 1923
Qy 1741 GCTTACAGCTGACGCGACCTGCCCATTTGGGAGACCAAGTCTATTCTGACGAGCGG 1800
Db 1924 GCTTACAGCTGACGCGACCTGCCCATTTGGGAGACCAAGTCTATTCTGACGAGCGG 1983
Qy 1801 CAGAGCCAGTGAAGAGCTGACTCTGCGGGGGAGGCTGAGCTGAAGAGAGCCCTTTGAA 1860
Db 1984 CAGAGCCAGTGAAGAGCTGACTCTGCGGGGAGGCTGAGCTGAAGAGAGCCCTTTGAA 2043
Qy 1981 AAGCATTTAAACGAGAGCTGCAATGGAATTTGAGAGAGCATCATGTCAAGAAC 1920
Db 2044 AAGCATTTAAACGAGAGCTGCAATGGAATTTGAGAGAGCATCATGTCAAGAAC 2103
Qy 1921 AGGTACCGGAAAGAGCTGGGAAAGTGGGAGTCAAGTCTATCTTTTGGGACAGATGAA 1980
Db 2104 AGGTACCGGAAAGAGCTGGGAAAGTGGGAGTCAAGTCTATCTTTTGGGACAGATGAA 2163
Qy 1981 ATCATTTAGGTCTCC 1995
Db 2164 ATCATTTAGGTCTCC 2178

RESULT 9
ABV20833
ID ABV20833 standard; cDNA; 5145 BP.
XX
AC ABV20833;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 20824.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
FN W02001.60860-A2.
PD
PD 23-AUG-2001.
XX
XX 20-FEB-2001; 2001WO-US005171.
XX
XX 17-FEB-2000; 2000US-0183319P.
XX 16-MAR-2000; 2000US-0189862P.
XX 25-MAY-2000; 2000US-0207454P.
XX 09-JUN-2000; 2000US-0211314P.
XX 18-JUL-2000; 2000US-0219007P.
XX 13-DEC-2000; 2000US-0255281P.
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PI Schlegel R, Endege WO, Monahan JB;
XX
XX WPI; 2001-662795/76.
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
XX prostate cells and correlating with presence of prostate cancer, useful
XX for detecting presence of prostate cancer, stage of prostate cancer.
PS Claim 1; Page 3419; 11750bp; English.

Db 2449 AAGCAGTTAAACGAGAGCTGCCAATGAAATTGGAGAGAGCATCATGTCAGAAC 2508
QY 1921 AGTCACGGGAAGAGCTGGGGAAGATGGGCACTCAGTTACCTTTTGGGCGACATGAA 1980
Db 2509 AGGTACGGGAAGAGCTGGGGAAGATGGGCACTCAGTTACCTTTTGGGCGACATGAA 2568
QY 1981 ATCATTTAGGTCTCC 1995
Db 2569 ATCATTTAGGTCTCC 2583

RESULT 10
ABV21080
ID ABV21080 standard; cDNA; 5145 BP.
AC ABV21080;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 21071.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX pharmacogenomic marker; gene; ss.
XX Homo sapiens.
XX MO200160860-A2.
XX
XX 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-0183319P.
XX 16-MAR-2000; 2000US-0189862P.
XX 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PI Schlegel R, Endege WO, Monahan UE,
XX
XX WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer. useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
XX
PS Claim 1; Page 3481; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the efficacy
CC of progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
XX Sequence 5145 BP; 1408 A; 1135 C; 1253 G; 1346 T; 0 U; 3 Other;

Query Match 94.9%; Score 1893; DB 5; Length 5145;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCCATGAGATGATGGAATCAATTTGTACTGAGAGTTGGTGGCTTCGTGAA 60
Db 589 ATGGCCATGAGATGATGGAATCAATTTGTACTGAGAGTTGGTGGCTTCGTGAA 648

QY 61 AGTGAAACGAAAAAGTGTGCTAAATGATAGCCGGCACTTTGTGAAATACATATCC 120
Db 649 AGTGAAACGAAAAAGTGTGCTAAATGATAGCCGGCACTTTGTGAAATACATATCC 708
QY 121 CACATTTTGGAAAGCCATTATATCACTGCTCCAAAGCTTTATGAAAGCGAAGTTGCAACG 180
Db 709 CACATTTTGGAAAGCCATTATATCACTGCTCCAAAGCTTTATGAAAGCGAAGTTGCAACG 768
QY 181 GACAAAGTGTATTTACAGAGCTCATCCAGATTGAGGAAACATPAGGTGACATTAT 240
Db 769 GACAAAGTGTATTTACAGAGCTCATCCAGATTGAGGAAACATPAGGTGACATTAT 828
QY 241 TGCAGTCAAGAGTTGATGTTACATCAATCAAGTCCCAAGATGTTGCTCTCTCTCA 300
Db 829 TGCAGTCAAGAGTTGATGTTACATCAAGTCCCAAGATGTTGCTCTCTCTCA 888
QY 301 GACTGTTTTCACCTGTACTTCTGGGTAACTGAGAAAGCTTCAACTCTGTTACCTG 360
Db 889 GACTGTTTTCACCTGTACTTCTGGGTAACTGAGAAAGCTTCAACTCTGTTACCTG 948
QY 361 CTGCAAGTGGGTTTGTGATGTTCTGCTGTTTCCCTGGGCTCTGTGAAAGAAATCC 420
Db 949 CTGCAAGTGGGTTTGTGATGTTCTGCTGTTTCCCTGGGCTCTGTGAAAGAAATCC 1008
QY 421 ACTGTAAGTCCCTACCTGATTTCTCAGCCTTGTACCTGTGCAACATTTGGCAACC 480
Db 1009 ACTGTAAGTCCCTACCTGATTTCTCAGCCTTGTACCTGTGCAACATTTGGCAACC 1068
QY 481 CGAATTCCTCCCAATCTTTATCTTGGGTCAGGAGAAATCCCTCAACAGAGCTGATA 540
Db 1069 CGAATTCCTCCCAATCTTTATCTTGGGTCAGGAGAAATCCCTCAACAGAGCTGATA 1128
QY 541 CAGAGAAATGGATGTTATGTTTAAATGCAAGTATACCTGTCAAGACCTGACTT 600
Db 1129 CAGAGAAATGGATGTTATGTTTAAATGCAAGTATACCTGTCAAGACCTGACTT 1188
QY 601 ATCCCCGAGTCTCACTTTCCTGCTGCTGTGATGACAGCTTTTGTGAAATTTTG 660
Db 1189 ATCCCCGAGTCTCACTTTCCTGCTGCTGTGATGACAGCTTTTGTGAAATTTTG 1248
QY 661 CCGTGTGTAACAATATGATGATTTTCAATGAAAGCAAAAGCCCTCAATGATGTT 720
Db 1249 CCGTGTGTAACAATATGATGATTTTCAATGAAAGCAAAAGCCCTCAATGATGTT 1308
QY 721 CTAGTCACTGTTTAACTGATGATCTCCGCTCCGCAACATCTGATGCTATCATATG 780
Db 1309 CTAGTCACTGTTTAACTGATGATCTCCGCTCCGCAACATCTGATGCTATCATATG 1368
QY 781 AAGAGATGACATGCTTTTAAATGATGATGATTTGTGAAAGAAAGAAAGAAAGCTACT 840
Db 1369 AAGAGATGACATGCTTTTAAATGATGATGATTTGTGAAAGAAAGAAAGAAAGCTACT 1428
QY 841 ATATCTCAAACTTCAATTTTCTGGGCAATCTCTGATGATGATGATGATGATGATG 900
Db 1429 ATATCTCAAACTTCAATTTTCTGGGCAATCTCTGATGATGATGATGATGATGATG 1488
QY 901 CAACTGAGAGCATCAGGCGCAAGCAAACTCAAGCTCTGACCTGAGAAAGCAAT 960
Db 1489 CAACTGAGAGCATCAGGCGCAAGCAAACTCAAGCTCTGACCTGAGAAAGCAAT 1548
QY 961 GAACCTGCTCCCTGCTCTCAAGAGGTGAGAAAGCAAGCAAGCAAGCAAGCAAGCA 1020
Db 1549 GAACCTGCTCCCTGCTCTCAAGAGGTGAGAAAGCAAGCAAGCAAGCAAGCAAGCA 1608
QY 1021 TGTGCGACTGTCTACTCAAGAGCAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1080
Db 1609 TGTGCGACTGTCTACTCAAGAGCAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1668
QY 1081 CCAAGCTGCTCCAGGCTGAGCGCTGCTGTTAAGAGCAAGCGCTGTTAAGAGCGCTC 1140
Db 1669 CCAAGCTGCTCCAGGCTGAGCGCTGCTGTTAAGAGCAAGCGCTGTTAAGAGCGCTC 1728

Db 949 CTTCAGAGTGGGTTTGCTGAGTTCTCTGTTGTTTCCCTGGGCTCTGTGAAGAAATCC 1008
Qy 421 ACTTAGTCCCTACCTGATCTTCTCAGCCTTGCTTACTTGTCGCAATTTGGGCCAAC 480
Db 1009 ACTTAATCTCCCTACCTGATCTTCTCAGCCTTGCTTACTTGTCGCAATTTGGGCCAAC 1068
Qy 481 CGAATTTCTCCCAATCTTTAATCTTGTCGTCAGAGATGTCTCAACAAGAGCTGATA 540
Db 1069 CGAATTTCTCCCAATCTTTAATCTTGTCGTCAGAGATGTCTCAACAAGAGCTGATA 1128
Qy 541 CAGAGAAATGGAGTGGTTATGTTTAAATGCCAGCTATACCTGTCCAAAGCTGACTT 600
Db 1129 CAGAGAAATGGAGTGGTTATGTTTAAATGCCAGCAATACCTGTCCAAAGCTGACTT 1188
Qy 601 ATCCCGAGTCTGATTTCTGCGTGTGCTGTGAATGACAGCTTTTGTGAGAAATTTTG 660
Db 1189 ATCCCGAGTCTGATTTCTGCGTGTGCTGTGAATGACAGCTTTTGTGAGAAATTTTG 1248
Qy 661 CCGTGGTTGGCAAAATCAGTAGATTTTCAATTGAGAAAGCAAAAGCTCCAAATGATGTT 720
Db 1249 CCGTGGTTGGCAAAATCAGTAGATTTTCAATTGAGAAAGCAAAAGCTCCAAATGATGTT 1308
Qy 721 CTAGTGACCTGTTAGTGGGATCTCCGCTCCGCCACCATGCTTATCGCTTACATCATG 780
Db 1309 CTAGTGACCTGTTAGTGGGATCTCCGCTCCGCCACCATGCTTATCGCTTACATCATG 1368
Qy 781 AAGAGATGACATGCTTTAGATGAGCTTACAGATTTGTGAAGAAAGAAAGCACTACT 840
Db 1369 AAGAGATGACATGCTTTAGATGAGCTTACAGATTTGTGAAGAAAGAAAGCACTACT 1428
Qy 841 ATATCTCCAACTTCAATTTTCTGGGCCAACTCTGACTATGAGAAAGATTTAAGAAC 900
Db 1429 ATATCTCCAACTTCAATTTTCTGGGCCAACTCTGACTATGAGAAAGATTTAAGAAC 1488
Qy 901 CAGACTGGAGCATCAGGGCCAAAGAGCAAACTCAAGCTGACCTGAGAAAGCCAAAT 960
Db 1489 CAGACTGGAGCATCAGGGCCAAAGAGCAAACTCAAGCTGACCTGAGAAAGCCAAAT 1548
Qy 961 GAACTGTCTCTGCTGTCTCAGAGGGTGGACAGAAAGAGAGAGCGCCCTCAGTCCACC 1020
Db 1549 GAACTGTCTCTGCTGTCTCAGAGGGTGGACAGAAAGAGAGAGCGCCCTCAGTCCACC 1608
Qy 1021 TGTGCGGACTCTGCTAATCTCAGAGGCAAGAGCAAAAGGCCGTGATCCGCCAGCGTG 1080
Db 1609 TGTGCGGACTCTGCTAATCTCAGAGGCAAGAGCAAAAGGCCGTGATCCGCCAGCGTG 1668
Qy 1081 CCCAGGCTGCCAGGCTGAGCGCTGTGTTAGAGGACAGAGCGCCCTGTTAAGCGCTC 1140
Db 1669 CCCAGGCTGCCAGGCTGAGCGCTGTGTTAGAGGACAGAGCGCCCTGTTAAGCGCTC 1728
Qy 1141 AGTGGGCTGCACCTGTCCGACAGAGGCTGAGAGCAGCATTAAGCTTCTCTTC 1200
Db 1729 AGTGGGCTGCACCTGTCCGACAGAGGCTGAGAGCAGCATTAAGCTTCTCTTC 1788
Qy 1201 TCTCTGATATCAATCAAGTTCATATTGACCCAGATGGCAGCATCTTTCATGAGCTTC 1260
Db 1789 TCTCTGATATCAATCAAGTTCATATTGACCCAGATGGCAGCATCTTTCATGAGCTTC 1848
Qy 1261 TCCCTATCAGAGATGCTTTGGAATCTAACAACCTTCCACTACCTGAGAGGAGCCAAC 1320
Db 1849 TCCCTATCAGAGATGCTTTGGAATCTAACAACCTTCCACTACCTGAGAGGAGCCAAC 1908
Qy 1321 AAGCTATGCAAGTCTTCCCTGTGTAGAGAACTATGAGAGAGACTCCGAAACAGTCTCT 1380
Db 1909 AAGCTATGCAAGTCTTCCCTGTGTAGAGAACTATGAGAGAGACTCCGAAACAGTCTCT 1968
Qy 1381 GATTAAGAGAGAGCCAGCATCCCAAGAACTGCAAGCCGCCAGGCTTTCAGACAGCCAG 1440
Db 1969 GATTAAGAGAGAGCCAGCATCCCAAGAACTGCAAGCCGCCAGGCTTTCAGACAGCCAG 2028
Qy 1441 AGCAAGCATTTGCTTGGTGTAGAACCAAGCAGAGCAGTGGCAAGGCCAAGAGTCCCTTTA 1500
Db 2029 AGCAAGCATTTGCTTGGTGTAGAACCAAGCAGAGCAGTGGCAAGGCCAAGAGTCCCTTTA 2088

Qy 1501 TCTCCACTGATCGAAGTGGAGCGGTGAGAGCAATTAACACACCAAGCTTCTTTCCGC 1560
Db 2089 TCTCCACTGATCGAAGTGGAGCGGTGAGAGCAATTAACACACCAAGCTTCTTTCCGC 2148
Qy 1561 CTTTCACAGCCACAGACACTCAGAGAGTGTGTGGCTGGGCTTAAGGGCTGGGAC 1620
Db 2149 CTTTCACAGCCACAGACACTCAGAGAGTGTGTGGCTGGGCTTAAGGGCTGGGAC 2208
Qy 1621 TCGAATATCTTGGCCCCCAAGCTTACCCCTTCCCTGACACAGCAGCTGTATTTTGC 1680
Db 2209 TCGAATATCTTGGCCCCCAAGCTTACCCCTTCCCTGACACAGCAGCTGTATTTTGC 2268
Qy 1681 ACAGAGTCTTACATTTCTGCTGCTGACAGCATCTAAGAGAGAGCCAGTACTCT 1740
Db 2269 ACAGAGTCTTACATTTCTGCTGCTGACAGCATCTAAGAGAGAGCCAGTACTCT 2328
Qy 1741 GCTTACAGCTGACAGCAGCTGCCACTTGGCGGAGCAAGCTTATCTGTGCGAGGCG 1800
Db 2329 GCTTACAGCTGACAGCAGCTGCCACTTGGCGGAGCAAGCTTATCTGTGCGAGGCG 2388
Qy 1801 CAGAGCCAAAGTGCAGAGCTGACTGTGGCGGAGCTGGCATGAAGAGAGCCCTTTGAA 1860
Db 2389 CAGAGCCAAAGTGCAGAGCTGACTGTGGCGGAGCTGGCATGAAGAGAGCCCTTTGAA 2448
Qy 1861 AAGCATTTAAACGAGAAAGCTGCAAAATGGAATTTGAGAGACATCATGTGAGAAC 1920
Db 2449 AAGCATTTAAACGAGAAAGCTGCAAAATGGAATTTGAGAGACATCATGTGAGAAC 2508
Qy 1921 AGGTACCGGAGAGAGCTGGGAAAGTGGGAGTCAAGTCTTTCGGGAGCATGGAA 1980
Db 2509 AGGTACCGGAGAGAGCTGGGAAAGTGGGAGTCAAGTCTTTCGGGAGCATGGAA 2568
Qy 1981 ATCATTTAGGCTCTCC 1995
Db 2569 ATCATTTAGGCTCTCC 2583

RESULT 12
ABV20978
ID ABV20978 standard; cDNA; 5145 BP.
XX
AC ABV20978;
XX
AC
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 20969.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN W0200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-0183319P.
XX 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 16-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
XX WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of

PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.

XX Claim 1, Page 3451, 11750pp, English.

XX The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV0010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC / (I) is also useful as a pharmacodynamic or pharmacogenomic marker
XX

SO Sequence 5145 BP; 1408 A; 1135 C; 1253 G; 1346 T; 0 U; 3 Other;

Query Match 94.9%; Score 1893; DB 5; Length 5145;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy	1	ATGGCCCATGAGATGATGAGACCTCAATTTGTAAGAGAGTGGTGGCTGTGGAGAA	60
Db	589	ATGGCCCATGAGATGATGAGACCTCAATTTGTAAGAGAGTGGTGGCTGTGGAGAA	648
Qy	61	AGTGAACGGAAGAGTGTCTAATGATAGCGGSCATTTGTGAATACAAATACATCC	120
Db	649	AGTGAACGGAAGAGTGTCTAATGATAGCGGSCATTTGTGAATACAAATACATCC	708
Qy	121	CACATTTTGAAGCCATTAATATCAACTGCTCCAGCTTATGAAGCGAAGTTGCAACAG	180
Db	709	CACATTTTGAAGCCATTAATATCAACTGCTCCAGCTTATGAAGCGAAGTTGCAACAG	768
Qy	181	GACAAAGTGTATTAATAGAGCTCATCCAGCATTCAGCGAAACATAAGTTGACATGAT	240
Db	769	GACAAAGTGTATTAATAGAGCTCATCCAGCATTCAGCGAAACATAAGTTGACATGAT	828
Qy	241	TGCAGTCAGAAAGTGTAGTTTACGATCAAGCTCCCAAGATGTGCTCTCTCTCA	300
Db	829	TGCAGTCAGAAAGTGTAGTTTACGATCAAGCTCCCAAGATGTGCTCTCTCTCA	888
Qy	301	GACTGTTTCTCACTGACTTCTGGGTAACGTGAGAAAGCTTCAACTCTGTTACCTG	360
Db	889	GACTGTTTCTCACTGACTTCTGGGTAACGTGAGAAAGCTTCAACTCTGTTACCTG	948
Qy	361	CTTGCAAGTGGTGTGTGAGTTCTCTCGTTGTTTCCCTGGCTCTGTGAAAGAAATCC	420
Db	949	CTTGCAAGTGGTGTGTGAGTTCTCTCGTTGTTTCCCTGGCTCTGTGAAAGAAATCC	1008
Qy	421	ACTCTAGTCCCTACCTGATCTTCTCAGCTTGTCTTACCTGTGCAACATTTGGGCCAAC	480
Db	1009	ACTCTAGTCCCTACCTGATCTTCTCAGCTTGTCTTACCTGTGCAACATTTGGGCCAAC	1068
Qy	481	CGAATCTTCCCAATCTTATCTTGTGCTGCAAGAGATGTCTCAACAAAGAGCTGATA	540
Db	1069	CGAATCTTCCCAATCTTATCTTGTGCTGCAAGAGATGTCTCAACAAAGAGCTGATA	1128
Qy	541	CAGCAAGATGGATGGTTATGTGTAAATGCGCAGCTATACCTGTCCAAAGGCTGACTTT	600
Db	1129	CAGCAAGATGGATGGTTATGTGTAAATGCGCAGCTATACCTGTCCAAAGGCTGACTTT	1188
Qy	601	ATCCCCGAGTCTCATTTCTGCGGTGCTGTGATGACAGCTTTTGTGAGAAATTTTG	660
Db	1189	ATCCCCGAGTCTCATTTCTGCGGTGCTGTGATGACAGCTTTTGTGAGAAATTTTG	1248
Qy	661	CCGTGTGTGACAAATCAGTAGATTTTATGAGAAAGCAAAAGCTTCAATGAGTGT	720
Db	1249	CCGTGTGTGACAAATCAGTAGATTTTATGAGAAAGCAAAAGCTTCAATGAGTGT	1308
Qy	721	CTAGTGCATGTTAGTGGATCTCCCGCTCCGCCACATCGCTATCGCTACATCATG	780

Db	1309	CTAGTGCATGTTAGTGGATCTCCCGCTCCGCCACATCGCTATCGCTACATCATG	1368
Qy	781	AAGAGATGACATGTCTTATGATGAGCTTACAGATTGTGAAAGAAAGAAAGCACTACT	840
Db	1369	AAGAGATGACATGTCTTATGATGAGCTTACAGATTGTGAAAGAAAGAAAGCACTACT	1428
Qy	841	ATATCTCCCAACTTCAATTTTCTGGGCAACTCTCTGACATATAGAAAGATTAAGAAC	900
Db	1429	ATATCTCCCAACTTCAATTTTCTGGGCAACTCTCTGACATATAGAAAGATTAAGAAC	1488
Qy	901	CAGCTGAGCATCAGGCGCAAGCAAAAGCAAACTGAGCTGTGACCTGAGAAAGCAAT	960
Db	1489	CAGCTGAGCATCAGGCGCAAGCAAAAGCAAACTGAGCTGTGACCTGAGAAAGCAAT	1548
Qy	961	GAACTGTCTCTGCTGTCTTCAAGAGGTGACAGAAAGCGAGACGCCCTCACTCAACC	1020
Db	1549	GAACTGTCTCTGCTGTCTTCAAGAGGTGACAGAAAGCGAGACGCCCTCACTCAACC	1608
Qy	1021	TGTGCGGACTGTGCTTACCTTACAGAGCAGCAGCAAAAGGCGCTGATCCCGCAGCTG	1080
Db	1609	TGTGCGGACTGTGCTTACCTTACAGAGCAGCAGCAAAAGGCGCTGATCCCGCAGCTG	1668
Qy	1081	CCGAGGCTGCCAGCGTGCAGCCGTGCTGTGAGAGACAGCCGCTGATCAGGCGCTC	1140
Db	1669	CCGAGGCTGCCAGCGTGCAGCCGTGCTGTGAGAGACAGCCGCTGATCAGGCGCTC	1728
Qy	1141	AGTGGCTGCACTGTCTCCGACAGCAGCTGGAAGACAGCAATTAAGCTCAAGCTCTTC	1200
Db	1729	AGTGGCTGCACTGTCTCCGACAGCAGCTGGAAGACAGCAATTAAGCTCAAGCTCTTC	1788
Qy	1201	TCTCTGATATCAATCAGTTTATATTTACAGCCAGCATGAGCATCTTACATGTGCTTC	1260
Db	1789	TCTCTGATATCAATCAGTTTATATTTACAGCCAGCATGAGCATCTTACATGTGCTTC	1848
Qy	1261	TCTCTATCAGAAAGATCTTTGAAATCTAACAACCTTCCACTCTGTGATGGAACCAAC	1320
Db	1849	TCTCTATCAGAAAGATCTTTGAAATCTAACAACCTTCCACTCTGTGATGGAACCAAC	1908
Qy	1321	AGCTATGCGAGTTCTCCCTGTTCAGGAATATCGAGAGCATCCCGAAACAGTCTCT	1380
Db	1909	AGCTATGCGAGTTCTCCCTGTTCAGGAATATCGAGAGCATCCCGAAACAGTCTCT	1968
Qy	1381	GATTAAGAGAGAGCCAGCATCTCCCAAGAAAGCTGCAACCGGCTTACAGACAGCAG	1440
Db	1969	GATTAAGAGAGAGCCAGCATCTCCCAAGAAAGCTGCAACCGGCTTACAGACAGCAG	2028
Qy	1441	AGCAAGCAATGATTCGATCGAGCAACAGCAGAGTGGACCGCCAGAGGTCCTTTTA	1500
Db	2029	AGCAAGCAATGATTCGATCGAGCAACAGCAGAGTGGACCGCCAGAGGTCCTTTTA	2088
Qy	1501	TCTCACTGATGAGAGTGGAGACGTGAGAGCAATTAACACACAGCTTCTTTTCGCG	1560
Db	2089	TCTCACTGATGAGAGTGGAGACGTGAGAGCAATTAACACACAGCTTCTTTTCGCG	2148
Qy	1561	CTTTCAACAGCAGCAGACCTCTCAGAAAGTGTGCTGCTGGGCTTAAAGGCTGGCAC	1620
Db	2149	CTTTCAACAGCAGCAGACCTCTCAGAAAGTGTGCTGCTGGGCTTAAAGGCTGGCAC	2208
Qy	1621	TGGATATCTTGGGCCCCGAGACCTTACCCCTTCCCTGACAGCAGCTGTATTTTGGC	1680
Db	2209	TGGATATCTTGGGCCCCGAGACCTTACCCCTTCCCTGACAGCAGCTGTATTTTGGC	2268
Qy	1681	ACAGATCTCTCACTTCTACTCTGAGCCATGACGAGAGCAGTGCAGATTACTCT	1740
Db	2269	ACAGATCTCTCACTTCTACTCTGAGCCATGACGAGAGCAGTGCAGATTACTCT	2328
Qy	1741	GCTTACAGCTGACAGCAGCTGCTGCTGAGAGCAAAAGTCTATTTCTGTGCGCAGCGG	1800
Db	2329	GCTTACAGCTGACAGCAGCTGCTGCTGAGAGCAAAAGTCTATTTCTGTGCGCAGCGG	2388
Qy	1801	CAGAGCAAGTGAACAGAGCTGACGCGGCGAGCTGCGATGAAGAGGCCCTTTGAA	1860

Db	2389	CAGAAGCCAAGTGA	CAGAGCTGACTCGCGGGGAGCTGGCATGAAAGAGGCCCTTTGAA	2448	
QY	1861	AAGCAGTTTAAACG	CAAGAGCTGC	CAATGTGAATTGGAGAGACATCATGTCCAGAGAAC	1920
Db	2449	AAGCAGTTTAAACG	CAGAGCTGC	CAATGTGAATTGGAGAGACATCATGTCCAGAGAAC	2508
QY	1921	AGGTCACGGGGAAGAGCTGGGG	AAAAGTGGG	CAGTCACTTTCAGCTTTTGGGGCAGCATGAAA	1980
Db	2509	AGGTCACGGGGAAGAGCTGGGG	AAAAGTGGG	CAGTCACTTTCAGCTTTTGGGGCAGCATGAAA	2568
QY	1981	ATCATTTAGAGTCTCC			
Db	2569	ATCATTTAGAGTCTCC	2583		

RESULT 13
ABV21092
ID ABV21092 standard; cDNA; 5145 BP
vv

DT 13-SEP-2002 (first entry)
 YY

Human prostate expression marker cDNA 21083.

KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
pharmacogenomic marker; gene; ss.

Homo sapiens

PN WO200160860-A2.

PD 23-AUG-2001.

PF 20-FEB-2001; 2001WO-US005171.

PR 17-FEB-2000; 2000US-0183319P

25-MAY-2000; 2000US-020745AP, PR

18-JUL-2000; 2000US-0219007P

13-DEC-2000; 200005-023528LP.
XX
XX

(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

Pf schlegel R, Endege WO, Monahan JE, XX

DK WPI; 2001-662795/76.
XX

PT Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer.

PS Claim 1; Page 3485; 11750pp; English.
v

The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement (II) is useful for: (a) assessing whether a patient is afflicted with prostate cancer; (b) monitoring the progression of prostate cancer in a patient; (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient; (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient; (e) selecting a composition for inhibiting prostate cancer in a patient; (f) assessing the prostate cell carcinogenic potential of a compound; (g) determining whether prostate cancer has metastasized in a patient; (h) assessing the aggressiveness or indolence of prostate cancer in a patient; (i) is also useful as a pharmacodynamic or pharmacogenomic marker

Sequence 5145 BP, 1408 A, 1135 C, 1253 G, 1346 T, 0 U, 3 Other;

Query Match	94.9%	Score 1893;	DB 5;	Length 5145;
Best Local Similarity	99.9%	Pred. No. 0;		
Matches 1993; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

QY	1	ATGSCCATGATGATGGAACTCAATATGTTACTAGAGGTTGGGCGCTGCTGGAA	60
Db	589	ATGGCCCATGAGATGATTTGAACTCAATTTGTTACTAGAGGTTGGGCGCTGCTGGAA	648
QY	61	AGTGGAA CGSAAAAAGTCTGCTTAATTGATAGCCGGCCATTGTTGGAAATCAATACATCC	120
Db	649	AGTGGAA CGSAAAAAGTCTGCTTAATTGATAGCCGGCCATTGTTGGAAATCAATACATCC	708
QY	121	CACATTTTGGAAAGCCATTAAATCAATGCTCCAGCTTATGAAAGGAAGTTGGAA CAG	180
Db	709	CACATTTTGGAAAGCCATTAAATCAATGCTCCAGCTTATGAAAGGAAGTTGGAA CAG	768
QY	181	GACAAAGTGTAAATTACAGAGCTTCATCAGCATTTACGCGAAACATAAGTTGACATTTGAT	240
Db	769	GACAAAGTGTAAATTACAGAGCTTCATCAGCATTTACGCGAAACATAAGTTGACATTTGAT	828
QY	241	TGCAGTCAGAAAGGTTTGATTTACAGATCAAAAGCTCCCAAGATGTTGCCCTCTCTTTCA	300
Db	829	TGCAGTCAGAAAGGTTTGATTTACAGATCAAAAGCTCCCAAGATGTTGCCCTCTCTTTCA	888
QY	301	GACTGTTTTCTCACTGACTCTTGCGGGTAACTGAGAAAGAGCTTCAACTCTGTTCACCTG	360
Db	889	GACTGTTTTCTCACTGACTCTTGCGGGTAACTGAGAAAGAGCTTCAACTCTGTTCACCTG	948
QY	361	CTTGACAGTGGGTTTGCTGAGTTCTCTCGTTGTTTCCCTGGCCTCTGTAAGAAAAATCC	420
Db	949	CTTGACAGTGGGTTTGCTGAGTTCTCTCGTTGTTTCCCTGGCCTCTGTAAGAAAAATCC	1008
QY	421	ACTGTAGTCCCTACCTGSCATTTCTCAGCCCTTGTTACCTGTTGCAACATTTGGGCCAAC	480
Db	1009	ACTGTAGTCCCTACCTGSCATTTCTCAGCCCTTGTTACCTGTTGCAACATTTGGGCCAAC	1068
QY	481	CGAATTTCTCCCAATCTTTATCTTTGGCTGCGACGAGATGTCCTCAACAGAGCTGATA	540
Db	1069	CGAATTTCTCCCAATCTTTATCTTTGGCTGCGACGAGATGTCCTCAACAGAGCTGATG	1128
QY	541	CAGAGAAATGGATTTGTTATTTGTTAAATGCCAGCTAATCTGTGCCAAAGCCTGACTTT	600
Db	1129	CAGAGAAATGGATTTGTTATTTGTTAAATGCCAGCAATACCTGTGCCAAAGCCTGACTTT	1188
QY	601	ATCCCCGAGCTCATTTTCTCGCGTGGCCGTGGAATGACAGCTTTTGTGAGAAAATTTTG	660
Db	1189	ATCCCCGAGCTCATTTTCTCGCGTGGCCGTGGAATGACAGCTTTTGTGAGAAAATTTTG	1248
QY	661	CGTGTGGACAAATCAGTAGATTTCAATTGAGAAAAGCCATTCATGATGTGTT	720
Db	1249	CGTGTGGACAAATCAGTAGATTTCAATTGAGAAAAGCCATTCATGATGTGTT	1308
QY	721	CTAGTGCACGTTTACTGTGGATCTCCGCGCTCCGACATGCGCTATGCGCTACATCATG	780
Db	1309	CTAGTGCACGTTTACTGTGGATCTCCGCGCTCCGACATGCGCTATGCGCTACATCATG	1368
QY	781	AAGAGATGACATGCTTTAGATGAAGTTTACATTTGTGTAAGAAAAAGACCTTACT	840
Db	1369	AAGAGATGACATGCTTTAGATGAAGTTTACATTTGTGTAAGAAAAAGACCTTACT	1428
QY	841	ATATCTCCAAATTTCAATTTTCTGGGCCAATCTCTGACATATGAGAAAGATTTAGAAC	900
Db	1429	ATATCTCCAAATTTCAATTTTCTGGGCCAATCTCTGACATATGAGAAAGATTTAGAAC	1488
QY	901	CAGACTGGAGCATCAGGGCCAAAGAGCAAACTCAAGCTGTGACCTGTGAGAAAGCCAAAT	960
Db	1489	CAGACTGGAGCATCAGGGCCAAAGAGCAAACTCAAGCTGTGACCTGTGAGAAAGCCAAAT	1548
QY	961	GAACTGTCTCTGCTGTCTCAGAGGGGTGACGAAAAACGACGCGCCCTCACTGTCACCC	1020
Db	1549	GAACTGTCTCTGCTGTCTCAGAGGGGTGACGAAAAACGACGCGCCCTCACTGTCACCC	1608
QY	1021	TGTGCGCATCTCTGTACTCACAAGGACAGACGAAAGCCCGGTGATCTCCGCCACGGTGG	1080
Db	1609	TGTGCGCATCTCTGTACTCACAAGGACAGACGAAAGCCCGGTGATCTCCGCCACGGTGG	1668

QY 1081 CCCAGCGTCCAGCCGTGACGCGCTGTTAGAGGACAGCCCGCTGTTAGAGCGCTC 1140
DB 1669 CCCAGCGTCCAGCCGTGACGCGCTGTTAGAGGACAGCCCGCTGTTAGAGCGCTC 1728
QY 1141 AGTGGCTGACCTGTCCGAGACAGGCTGGAAGACAGCAATTAAGCTCAAGCGTTCTTC 1200
DB 1729 AGTGGCTGACCTGTCCGAGACAGGCTGGAAGACAGCAATTAAGCTCAAGCGTTCTTC 1788
QY 1201 TCTCTGATATCAAAATCAAGTTTATATTCAGCCAGATGAGCAATCTTACATGCTTC 1260
DB 1789 TCTCTGATATCAAAATCAAGTTTATATTCAGCCAGATGAGCAATCTTACATGCTTC 1848
QY 1261 TCTCTATCAGAAAGATGTTGAAATCTAACAACCTTCCACTCTGATGAGGACCAAC 1320
DB 1849 TCTCTATCAGAAAGATGTTGAAATCTAACAACCTTCCACTCTGATGAGGACCAAC 1908
QY 1321 AAGCTATGCCAGTTCCTCCCTGTTACAGAACTATCGAGAGAGACTCCCGAAACAGTTCCT 1380
DB 1909 AAGCTATGCCAGTTCCTCCCTGTTACAGAACTATCGAGAGAGACTCCCGAAACAGTTCCT 1968
QY 1381 GATTAAGAGAGAACAGCATCCCAAGAGCTGACAGCCGCGCTTACAGACGACG 1440
DB 1969 GATTAAGAGAGAACAGCATCCCAAGAGCTGACAGCCGCGCTTACAGACGACG 2028
QY 1441 AGCAAGGATTTGATTCGGTCAAGACAGGAGAGTGGACCGCCAGAGGTCCCTTTA 1500
DB 2029 AGCAAGGATTTGATTCGGTCAAGACAGGAGAGTGGACCGCCAGAGGTCCCTTTA 2088
QY 1501 TCTCCACTGATGAGAGTGGAGAGCTGAGAGCAATTAACACACAGCTTCCTTTCCGCG 1560
DB 2089 TCTCCACTGATGAGAGTGGAGAGCTGAGAGCAATTAACACACAGCTTCCTTTCCGCG 2148
QY 1561 CTCTCCACAGCAGCAGCAGCTTCAGAGAGTCTGCTGCGCTGAGGCTTTAGGCGTGCAC 1620
DB 2149 CTCTCCACAGCAGCAGCAGCTTCAGAGAGTCTGCTGCGCTGAGGCTTTAGGCGTGCAC 2208
QY 1621 TCGGATATCTTGGCCCCCGAGACCTGACCTTCCCTGACAGAGCTGGTATTTTGGC 1680
DB 2209 TCGGATATCTTGGCCCCCGAGACCTGACCTTCCCTGACAGAGCTGGTATTTTGGC 2268
QY 1681 ACAGAGTCTCTACACTTCTACTGCTGCTCAGCAGCTTACGAGAGCAGTGCAGTACTCT 1740
DB 2269 ACAGAGTCTCTACACTTCTACTGCTGCTCAGCAGCTTACGAGAGCAGTGCAGTACTCT 2328
QY 1741 GCTTACAGCTGACGACGCTGCCCACTTGGCGAGACCAAGCTATTTCTGTGCGCAGCGG 1800
DB 2329 GCTTACAGCTGACGACGCTGCCCACTTGGCGAGACCAAGCTATTTCTGTGCGCAGCGG 2388
QY 1801 CAGAAGCCAGTGAAGAGCTGACTGCGCGGAGAGCTGGAGTGAAGAGCCCTTTGAA 1860
DB 2389 CAGAAGCCAGTGAAGAGCTGACTGCGCGGAGAGCTGGAGTGAAGAGCCCTTTGAA 2448
QY 1861 AAGCAGTTTAAAGCAGAAAGCTGCCAAATGGAATTTGAGAGAGCATCATGTGAGAAAC 1920
DB 2449 AAGCAGTTTAAAGCAGAAAGCTGCCAAATGGAATTTGAGAGAGCATCATGTGAGAAAC 2508
QY 1921 AGGTCAAGGAGAGAGCTGGGGAAGTGGGAGTCAAGTCTAGCTTTTGGGCGACATGAA 1980
DB 2509 AGGTCAAGGAGAGAGCTGGGGAAGTGGGAGTCAAGTCTAGCTTTTGGGCGACATGAA 2568
QY 1981 ATGATTGAGGTCTCC 1995
DB 2569 ATGATTGAGGTCTCC 2583

RESULT 14

ABV21312 standard; cDNA; 5145 BP.

ID ABV21312;
AC ABV21312;
DT 13-SEP-2002 (first entry)
XX

DE Human prostate expression marker cDNA 21303.
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW Pharmacogenomic marker; gene; ss.
XX Homo sapiens.
OS MO200160860-A2.
PN 23-AUG-2001.
PD 20-FEB-2001; 2001WO-US005171.
PF 17-FEB-2000; 2000US-0183119P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0213114P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX Schlegel R, Endege WO, Monahan JE;
PI MPI; 2001-662795/76.
DR
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS Claim 1; Page 3539; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (1) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (1) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (1) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
SQ Sequence 5145 BP; 1408 A; 1135 C; 1253 G; 1346 T; 0 U; 3 Other;
Query Match 94.9%; Score 1893; DB 5; Length 5145;
Best Local Similarity 99.9%; Pred No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGGCCCATGAGATGATTGGAATCAATTTGTTACTGAGAGGTGGCTGCTGAGAA 60
DB 589 ATGGCCCATGAGATGATTGGAATCAATTTGTTACTGAGAGGTGGCTGCTGAGAA 648
QY 61 AGTGAACGGAAAAAGTGTCTTAATTGATAGCCGGCCATTGTTGGAATTAATACATCC 120
DB 649 AGTGAACGGAAAAAGTGTCTTAATTGATAGCCGGCCATTGTTGGAATTAATACATCC 708
QY 121 CACATTTTGAAGCCATTATATCACTGCTCCAGCTTATGAGGAGGATTGCAACAG 180
DB 709 CACATTTTGAAGCCATTATATCACTGCTCCAGCTTATGAGGAGGATTGCAACAG 768
QY 181 GACAAAGTGTATTAACAGAGCTCATCCAGATTCAGCGAAACATTAAGTTGACATTGAT 240
DB 769 GACAAAGTGTATTAACAGAGCTCATCCAGATTCAGCGAAACATTAAGTTGACATTGAT 828
QY 241 TGCAGTCAGAAAGTGTATTAACAGATCAAACTCCCAAGATGTTGCCCTCTCTCTTCA 300
DB 829 TGCAGTCAGAAAGTGTATTAACAGATCAAACTCCCAAGATGTTGCCCTCTCTCTTCA 888
QY 301 GACTGTTTCTACGTACTTCTGGGTAACTGAGAAAGCTTCAACTCTGTTACCTG 360

Db	889	GACGTGTTTCTCACTGACTTCTTGGGTAACTGGAGAAAGCTTCAACTGTGTCACTGG	948
Qy	361	CTTGCAAGTGGGTTTGCTGAAGTTCTCTCGTTGTTTCCCTGGGCTCTGTGAAGAAATATCC	420
Db	949	CTTGCAAGTGGGTTTGCTGAAGTTCTCTCGTTGTTTCCCTGGGCTCTGTGAAGAAATATCC	1008
Qy	421	ACTGTAGTCCCTACCTGCAATTTCTGACGCTTGCTTACCTGTGGCAACATTGGGCAACC	480
Db	1009	ACTGTAGTCCCTACCTGCAATTTCTGACGCTTGCTTACCTGTGGCAACATTGGGCAACC	1068
Qy	481	CGAATTCCTCCCAATCTTTATCTTTGGCTGCACGACAGATGTCTCAACAAGAGCTGATA	540
Db	1069	CGAATTCCTCCCAATCTTTATCTTTGGCTGCACGACAGATGTCTCAACAAGAGCTGATA	1128
Qy	541	CACGACAAATGGGATTTGGTTATGTGTAAATGCCAGCTATACCTGTCCAAAGCTGACTTT	600
Db	1129	CACGACAAATGGGATTTGGTTATGTGTAAATGCCAGCTATACCTGTCCAAAGCTGACTTT	1188
Qy	601	ATCCCGAGTCTCATTTCTCGGCTGTCCTGTGAATGACAGCTTTGTGAGAAAATTTTG	660
Db	1189	ATCCCGAGTCTCATTTCTCGGCTGTCCTGTGAATGACAGCTTTGTGAGAAAATTTTG	1248
Qy	661	CCGTGTGGACAAATCACTAATTTCAATTGAAAGCAAAAGCTCCAAATGATGTGT	720
Db	1249	CCGTGTGTGGACAAACATGTAAATTTCAATTGAAAGCAAAAGCTCCAAATGATGTGT	1308
Qy	721	CTAGTSCACTGTTTAGCTGGATCTCCCGCTCCGACACATGCGTATCGCTTAATGATG	780
Db	1309	CTAGTSCACTGTTTAGCTGGATCTCCCGCTCCGACACATGCGTATCGCTTAATGATG	1368
Qy	781	AAGAGATGGAACATGTCCTTTAGATGATGAAGCTTACAGATTTGTGAAAGAAAAGACTTACT	840
Db	1369	AAGAGATGGAACATGTCCTTTAGATGATGAAGCTTACAGATTTGTGAAAGAAAAGACTTACT	1428
Qy	841	ATATCTCCCAAACCTTCAATTTTCTGGGCCAACTCCCTGACCTATGGAAGAAATTAATAAC	900
Db	1429	ATATCTCCCAAACCTTCAATTTTCTGGGCCAACTCCCTGACCTATGGAAGAAATTAATAAC	1488
Qy	901	CAGACTGAGCATCAAGGCCAAAGACAAACCTCAAGCTGTGCACTTGGAAAGCCCAAT	960
Db	1489	CAGACTGAGCATCAAGGCCAAAGACAAACCTCAAGCTGTGCACTTGGAAAGCCCAAT	1548
Qy	961	GAACTGTCCTCCGCTGCTCAGAGGGTGGACAGAAAAGCGAGACGCCCTCACTCAACC	1020
Db	1549	GAACTGTCCTCCGCTGCTCAGAGGGTGGACAGAAAAGCGAGACGCCCTCACTCAACC	1608
Qy	1021	TGTGCGCACTCTGTCACTCTCAGAGCGACAGACAAAGGCCCTGCACTCCGCGACGGTG	1080
Db	1609	TGTGCGCACTCTGTCACTCTCAGAGCGACAGACAAAGGCCCTGCACTCCGCGACGGTG	1668
Qy	1081	CCGAGCTGGCCAGGCTCAGCGCTCAGCGTCTGTTAAGGACAGCCCGCTGTACAGGCGCTT	1140
Db	1669	CCGAGCTGGCCAGGCTCAGCGCTCAGCGTCTGTTAAGGACAGCCCGCTGTACAGGCGCTT	1728
Qy	1141	AGTGGGCTGCACCTGTCCGACAGACAGGCTGGAAACAGCAATTAAGCTCAAGCGTTCTTC	1200
Db	1729	AGTGGGCTGCACCTGTCCGACAGACAGGCTGGAAACAGCAATTAAGCTCAAGCGTTCTTC	1788
Qy	1201	TCTCTGGATATCAAAATCAGTTTCATTTACGCGCAGATGGCAGACATCTTACATGGCTTC	1260
Db	1789	TCTCTGGATATCAAAATCAGTTTCATTTACGCGCAGATGGCAGACATCTTACATGGCTTC	1848
Qy	1261	TCTCTCATAGAAAGATGGCTTGGAAATACCAAAACCTTCCACTCTCTGATATGGGACCAAC	1320
Db	1849	TCTCTCATAGAAAGATGGCTTGGAAATACCAAAACCTTCCACTCTCTGATATGGGACCAAC	1908
Qy	1321	AAGCTATGCCAGTTCTCCCTGTTCAAGAACTATCGAGGACGACTCCGCAACCACTGCT	1380
Db	1909	AAGCTATGCCAGTTCTCCCTGTTCAAGAACTATCGAGGACGACTCCGCAACCACTGCT	1968
Qy	1381	GATTAAGAGAGAGCAGCATCCCAAAACCTGCAACCCGCGCTTCAAGACCCAG	1440
Db	1969	GATTAAGAGAGAGCAGCATCCCAAAACCTGCAACCCGCGCTTCAAGACCCAG	2028

QY	1441	AGCAAGCATTTGCAATTCCGTCAGAAACGACGAGTGGCACC GCCCAAGATCCCTTTTA	1500
Db	2029	AGCAACGATTTGCTATTCGGTCAGAACGACGAGTGGCACC GCCCAAGATCCCTTTTA	2088
QY	1501	TCCTCACTCATCGAAGTGGAGGCGTGGAGGACATTAACAACACAGCTTCTTTTCGGC	1560
Db	2089	TCCTCACTCATCGAAGTGGAGGCGTGGAGGACATTAACAACAACAGCTTCTTTTCGGC	2148
QY	1561	CTTTCCACGACGACAGACACTCAACGAAATCTGTGGCTGGGCGCTTAAGGGCTGGCAC	1620
Db	2149	CTTTCCACGACGACAGACACTCAACGAAATCTGTGGCTGGGCGCTTAAGGGCTGGCAC	2208
QY	1621	TCGGATATCTTTGGCCCCCCAGACCTCTACCCCTTCCCTGACACAGCACTGTATTTTGGCC	1680
Db	2209	TCGGATATCTTTGGCCCCCCAGACCTCTACCCCTTCCCTGACACAGCACTGTATTTTGGCC	2268
QY	1681	ACAGAGTCTCTCACTTTTACTCTTCCCTCAAGCCATCTACGAGGCGAGTGGCACTACTCT	1740
Db	2269	ACAGAGTCTCTCACTTTTACTCTTCCCTCAAGCCATCTACGAGGCGAGTGGCACTACTCT	2328
QY	1741	GCTTACAGCTGCAGCCAGCTGCCACTTTCGCGAGACCAAGTCTATTCTGGCGCAGCGG	1800
Db	2329	GCTTACAGCTGCAGCCAGCTGCCACTTTCGCGAGACCAAGTCTATTCTGGCGCAGCGG	2388
QY	1801	CAGAAAGCCAAAGTGACAGACTGACTCGCGGCGGAGCTGGCATTAAGAAGCCCTTTTGA	1860
Db	2389	CAGAAAGCCAAAGTGACAGACTGACTCGCGGCGGAGCTGGCATTAAGAAGCCCTTTTGA	2448
QY	1861	AAGCAGTTTAAACGACGAAAGCTGCCTCAATGGAATTTGGAGAGCATCTGTACAGAAAC	1920
Db	2449	AAGCAGTTTAAACGAGAAAGCTGCCTCAATGGAATTTGGAGAGCATCTGTACAGAAAC	2508
QY	1921	AGGTACCGGGAAAGAGCTGGGGGAAAGTGGGCAATGCACTTTCGGGCAAGATGAA	1980
Db	2509	AGGTACCGGGAAAGAGCTGGGGGAAAGTGGGCAATGCACTTTCGGGCAAGATGAA	2568
QY	1981	ATCATTGAGGTCTCC	1995
Db	2569	ATCATTGAGGTCTCC	2583
RESULT 15			
ID	ABV21316	ABV21316 standard; cDNA, 5145 BP.	
XX	AC	ABV21316;	
XX	DT	13-SEP-2002 (first entry)	
XX	DE	Human prostate expression marker cDNA 21307.	
XX	KW	Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;	
XX	XX	pharmacogenomic marker; gene; ss.	
OS	XX	Homo sapiens.	
XX	XX	WO200160860-A2.	
XX	PD	23-AUG-2001.	
XX	PF	20-FEB-2001; 2001WO-US005171.	
XX	PR	17-FEB-2000; 2000US-018319P.	
XX	PR	16-MAR-2000; 2000US-0189862P.	
XX	PR	25-MAY-2000; 2000US-0207454P.	
XX	PR	09-JUN-2000; 2000US-0211314P.	
XX	PR	18-JUL-2000; 2000US-0219007P.	
XX	XX	13-DEC-2000; 2000US-0255281P.	
XX	XX	(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.	
PI	XX	Schlegel R, Endege WO, Monahan JE:	

Db 2329 GCTTACAGCTGACGACCACTCCCACTTGGCGAGACCAAGTATCTGTGCGGAGCGG 2388
Qy 1801 CAGAAGCCAACTGACAGAGCTACCTCGGCGGAGCTGGCATGAGAGACCCCTTTGAA 1860
Db 2389 CAGAAGCCAACTGACAGAGCTACCTCGGCGGAGCTGGCATGAGAGACCCCTTTGAA 2448
Qy 1861 AAGCAGTTTAAACCCAGAAAGCTGCGCAATGGAATTTGGAGAGACATCATGCGAGAAC 1920
Db 2449 AAGCAGTTTAAACCCAGAAAGCTGCGCAATGGAATTTGGAGAGACATCATGCGAGAAC 2508
Qy 1921 AGGTCACGGGAAAGAGCTGGGAAAGTGCGCATGCTCTAGCTTTTGGGCGACATGAA 1980
Db 2509 AGGTCACGGGAAAGAGCTGGGAAAGTGCGCATGCTCTAGCTTTTGGGCGACATGAA 2568
Qy 1981 ATCATTTGAGGTCTCC 1995
Db 2569 ATCATTTGAGGTCTCC 2583

RESULT 16
ABV26826
ID ABV26826 standard; cDNA; 5145 BP.
XX
AC ABV26826;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 26817.
XX
KM Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX pharmacogenomic marker; gene; ss.
XX Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-0183119P.
XX
PR 16-MAR-2000; 2000US-0189862P.
XX
PR 25-MAY-2000; 2000US-0207454P.
XX
PR 09-JUN-2000; 2000US-0211314P.
XX
PR 18-JUL-2000; 2000US-0219007P.
XX
PR 13-DEC-2000; 2000US-0255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JB;
XX
XX WPI; 2001-662795/76.
XX
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
XX prostate cells and correlating with presence of prostate cancer. Useful
XX for detecting presence of prostate cancer, stage of prostate cancer.
XX
XX Claim 1; Page 5420-5421; 11750pp; English.

The invention relates to an isolated nucleic acid molecule (I) comprising
a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
specification or its complement. (I) is useful for: (a) assessing whether
a patient is afflicted with prostate cancer; (b) monitoring the
progression of prostate cancer in a patient; (c) assessing the efficacy
of a test compound to inhibit prostate cancer in a patient; (d) assessing
the efficacy of a therapy for inhibiting prostate cancer in a patient;
(e) selecting a composition for inhibiting prostate cancer in a patient;
(f) assessing the prostate cell carcinogenic potential of a compound; (g)
determining whether prostate cancer has metastasized in a patient; (h)
assessing the aggressiveness or indolence of prostate cancer in a patient
; (I) is also useful as a pharmacodynamic or pharmacogenomic marker

Sequence 5145 BP; 1408 A; 1135 C; 1253 G; 1346 T; 0 U; 3 Other;

Query Match 94.9%; Score 1893; DB 5; Length 5145;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGCCCATGAGATGATTTGGAATCTCAATTTGTACTGAGAGTTGGTCTCTGTGGA 60
Db 589 ATGGCCCATGAGATGATTTGGAATCTCAATTTGTACTGAGAGTTGGTCTCTGTGGA 648
Qy 61 AGTGAACGGAAAAAGTCTGCTAATTTGATAGCCGCCCATTTTGGAAATACAAATCATCC 120
Db 649 AGTGAACGGAAAAAGTCTGCTAATTTGATAGCCGCCCATTTTGGAAATACAAATCATCC 708
Qy 121 CACATTTTGAAGCCATTATATCACTGCTCCAAAGCTTTATGAAAGCGAGTTGCAACAG 180
Db 709 CACATTTTGAAGCCATTATATCACTGCTCCAAAGCTTTATGAAAGCGAGTTGCAACAG 768
Qy 181 GACAAAGTGTAAATTACAGAGCTCATCCAGATTCAGCGAAACATAAGGTGACATTTGAT 240
Db 769 GACAAAGTGTAAATTACAGAGCTCATCCAGATTCAGCGAAACATAAGGTGACATTTGAT 828
Qy 241 TGCAGTCAGAAAGTTGTAAGTTTACAGATCAAGATGTTGCTCTCTCTTCA 300
Db 829 TGCAGTCAGAAAGTTGTAAGTTTACAGATCAAGATGTTGCTCTCTCTTCA 888
Qy 301 GACTGTTTTCACCTGTAATCTTGGGTAAATCGAGAAAGAGCTTCAACTCTTCACTTG 360
Db 889 GACTGTTTTCACCTGTAATCTTGGGTAAATCGAGAAAGAGCTTCAACTCTTCACTTG 948
Qy 361 CTTCAGAGTGGGTTTGGCTGAGTTCTCTGTTGTTTCCCTGGGCTCTGTGAAAGAAATTC 420
Db 949 CTTCAGAGTGGGTTTGGCTGAGTTCTCTGTTGTTTCCCTGGGCTCTGTGAAAGAAATTC 1008
Qy 421 ACTTAAGTCCCTACCTGATTTCTACGCTTGTCTTACCTGTGCAACATTTGGGCAACC 480
Db 1009 ACTTAAGTCCCTACCTGATTTCTACGCTTGTCTTACCTGTGCAACATTTGGGCAACC 1068
Qy 481 CGAATTCCTCCCAATCTTTAATCTTGGGCTGCAAGGAAATGTCCTCAACAGAGCTATA 540
Db 1069 CGAATTCCTCCCAATCTTTAATCTTGGGCTGCAAGGAAATGTCCTCAACAGAGCTATA 1128
Qy 541 CAGCAGAAATGGGATTTGTTATGTTTAAATGCCAGTATACCTGTCCAAAGCTGACTTT 600
Db 1129 CAGCAGAAATGGGATTTGTTATGTTTAAATGCCAGTATACCTGTCCAAAGCTGACTTT 1188
Qy 601 ATCCCGAGTCTCAATTTCTGCGTGTGCTGTGAATGACAGCTTTTGTGAAAAATTTTG 660
Db 1189 ATCCCGAGTCTCAATTTCTGCGTGTGCTGTGAATGACAGCTTTTGTGAAAAATTTTG 1248
Qy 661 CCGTGTGTTGACAAATCAGTATGATTTCAATGAGAAAGCAAAAGCCCTCAATGATGTGT 720
Db 1249 CCGTGTGTTGACAAATCAGTATGATTTCAATGAGAAAGCAAAAGCCCTCAATGATGTGT 1308
Qy 721 CTAGTGCACTGTTTAACTGGGATCTCCGCTCCGCCACATCGCTATGCTTCATCATG 780
Db 1309 CTAGTGCACTGTTTAACTGGGATCTCCGCTCCGCCACATCGCTATGCTTCATCATG 1368
Qy 781 AAGAGATGGAACATGCTTTAATGAAAGCTTACAGATTTGTGAAAGAAAAAGCACTTACT 840
Db 1369 AAGAGATGGAACATGCTTTAATGAAAGCTTACAGATTTGTGAAAGAAAAAGCACTTACT 1428
Qy 841 ATATCTCCAAACTTCAATTTTCTGGGCCAAATCCCTGAGCATGAGAAAGAAATTAAGAC 900
Db 1429 ATATCTCCAAACTTCAATTTTCTGGGCCAAATCCCTGAGCATGAGAAAGAAATTAAGAC 1488
Qy 901 CAGACTGAGCATCAGGGCCAAAGAGCAAACTCAAGCTCTGACCTGAGAGAACCAAT 960
Db 1489 CAGACTGAGCATCAGGGCCAAAGAGCAAACTCAAGCTCTGACCTGAGAGAACCAAT 1548
Qy 961 GAACCTGTCTCTGCTCTCAAGAGGTGACAGAAAGCGAGACGCGCTCAGTCCACCC 1020
Db 1549 GAACCTGTCTCTGCTCTCAAGAGGTGACAGAAAGCGAGACGCGCTCAGTCCACCC 1608

QY 1021 TGTCCGACTCTGCTACTCTCAGAGGCGAGCAAAAGCCCGTGCATCCCGCAGCGTG 1080
DB 1606 TGTGCCGACTCTGCTACTCTCAGAGGCGAGCAAAAGCCCGTGCATCCCGCAGCGTG 1668
QY 1081 CCCAGCGTCCCGACCGTGCAGCCGTGCTGTAGAGAGCAAGCCCGTGTACAGCGCTC 1140
DB 1669 CCCAGCGTCCCGACCGTGCAGCCGTGCTGTAGAGAGCAAGCCCGTGTACAGCGCTC 1728
QY 1141 AGTGGGCTGCACTGTCCGAGACAGCTGGAAGACAGCAATTAAGCTCAAGCGTTCCTTC 1200
DB 1729 AGTGGGCTGCACTGTCCGAGACAGCTGGAAGACAGCAATTAAGCTCAAGCGTTCCTTC 1788
QY 1201 TCTCTGATATCAAAATCAGTTTCAATTCAGCCAGCATGCGAGCATCTTACATGCGCTC 1260
DB 1789 TCTCTGATATCAAAATCAGTTTCAATTCAGCCAGCATGCGAGCATCTTACATGCGCTC 1848
QY 1261 TCTCTATCAGAAAGATGTTTGAATTAATAAATCTTCACTCTGATGAGGACCAAC 1320
DB 1849 TCTCTATCAGAAAGATGTTTGAATTAATAAATCTTCACTCTGATGAGGACCAAC 1908
QY 1321 AAGCTATGCCAGTTTCCCTGTTTCAGAGAACTATGCGAGAGACTCCCGAAACAGTCTCT 1380
DB 1909 AAGCTATGCCAGTTTCCCTGTTTCAGAGAACTATGCGAGAGACTCCCGAAACAGTCTCT 1968
QY 1381 GATAAGAGAGAGCCAGCATCCCGAAGAGCTGAGACCGCGCAGAGCTTCAGACAGCAG 1440
DB 1969 GATAAGAGAGAGCCAGCATCCCGAAGAGCTGAGACCGCGCAGAGCTTCAGACAGCAG 2028
QY 1441 AGCAAGCGATTGCGATTCGATCAGAACCGAGAGAGTGGACCGCGCAGAGTCCCTTTTA 1500
DB 2029 AGCAAGCGATTGCGATTCGATCAGAACCGAGAGAGTGGACCGCGCAGAGTCCCTTTTA 2088
QY 1501 TCTCCACTGATGAGAGTGGAGAGCTGAGAGAACTTAACAACAGCTTCTTTTCGCG 1560
DB 2089 TCTCCACTGATGAGAGTGGAGAGCTGAGAGAACTTAACAACAGCTTCTTTTCGCG 2148
QY 1561 CTTTCCACGAGCCAGCAGCACTCAGCAAGTCTGTGAGCTGAGGCTTAAGGGCTGAGAC 1620
DB 2149 CTTTCCACGAGCCAGCAGCACTCAGCAAGTCTGTGAGCTGAGGCTTAAGGGCTGAGAC 2208
QY 1621 TCGGATATCTTGGCCCCCGCAGACCTTACCCCTTCCTGACAGAGCTGATATTTTGGC 1680
DB 2209 TCGGATATCTTGGCCCCCGCAGACCTTACCCCTTCCTGACAGAGCTGATATTTTGGC 2268
QY 1681 ACGAGATCTCTCAACTTCTACTCTGCTCAGCCATCTAGCGAGGAGCTGCCAATCTCT 1740
DB 2269 ACGAGATCTCTCAACTTCTACTCTGCTCAGCCATCTAGCGAGGAGCTGCCAATCTCT 2328
QY 1741 GCCTACAGCTGAGCCAGCTGCCCACTTGGGAGACCAAGTCTATTCTGTGGCGCAGGCG 1800
DB 2329 GCCTACAGCTGAGCCAGCTGCCCACTTGGGAGACCAAGTCTATTCTGTGGCGCAGGCG 2388
QY 1801 CAGAACCCAGTACAGAGCTGACTCGCGCGAGCTGATGAAGAGAGCCCTTTGAA 1860
DB 2389 CAGAACCCAGTACAGAGCTGACTCGCGCGAGCTGATGAAGAGAGCCCTTTGAA 2448
QY 1861 AAGCAGTTTAAAGCCAGAGCTGCCAATGAAATTTGAGAGAGCATCATGTCAGAGAAC 1920
DB 2449 AAGCAGTTTAAAGCCAGAGCTGCCAATGAAATTTGAGAGAGCATCATGTCAGAGAAC 2508
QY 1921 AGGTACCGGGAAAGAGCTGGGGAAAGTGGGAGCTCAGCTTCTTTCGGGCAAGAGGAA 1980
DB 2509 AGGTACCGGGAAAGAGCTGGGGAAAGTGGGAGCTCAGCTTCTTTCGGGCAAGAGGAA 2568
QY 1981 ATCATTGAGGTCTCC 1995
DB 2569 ATCATTGAGGTCTCC 2583

RESULT 17
ABV27131
ID ABV27131 standard; cDNA; 5145 BP.
XX

AC ABV27131;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 27122.
XX
DE Human prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KM pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN MO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-0183119P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
XX WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous strate of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer. Stage of prostate cancer.
XX
XX Claim 1; Page 5499-5500; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (i) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (i) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
SQ Sequence 5145 BP; 1408 A; 1135 C; 1253 G; 1346 T; 0 U; 3 Other;
Query Match 94.9%; Score 1893; DB 5; Length 5145;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGGCCCATGAGATGATTGGAACCTCAATTTGTTACTGAGAGTGTGCTGCTGGA 60
DB 589 ATGGCCCATGAGATGATTGGAACCTCAATTTGTTACTGAGAGTGTGCTGCTGGA 648
QY 61 AGTGAAGGAAAAGTGTGCTAAATGATGAGCGGCGCAATTTGGAATACATATCC 120
DB 649 AGTGAAGGAAAAGTGTGCTAAATGATGAGCGGCGCAATTTGGAATACATATCC 708
QY 121 CACATTTTGAAGCCATTATATCACTGCTCCAACTTATGAGGAAAGTTGCAACAG 180
DB 709 CACATTTTGAAGCCATTATATCACTGCTCCAACTTATGAGGAAAGTTGCAACAG 768
QY 181 GACAAAGTTTAAATTAACAGAGCTCATCCGATTCAGGCAACATTAAGTTGACATTAT 240
DB 769 GACAAAGTTTAAATTAACAGAGCTCATCCGATTCAGGCAACATTAAGTTGACATTAT 828
QY 241 TCGAGTCAGAGGTTGATTTACATCAAGCTCCCAAGATGTTGCTCTGCTTCA 300

Db 829 TCGACTCAGAAAGTTGATGTTACGATCAAAAGCTCCAGAGTGTGCTCTCTCTCA 888
Qy 301 GACTGTTTCTCACTGTAATTCGTGGTAACTGAGAGAGCTTCAATCTGTTCACTG 360
Db 889 GACTGTTTCTCACTGTAATTCGTGGTAACTGAGAGAGCTTCAATCTGTTCACTG 948
Qy 361 CTTCGAGGTGGGTTTGTGAGTCTCTGTTGTTTCCCTGGCTCTGTGAAGAAAATCC 420
Db 949 CTTCGAGGTGGGTTTGTGAGTCTCTGTTGTTTCCCTGGCTCTGTGAAGAAAATCC 1008
Qy 421 ACTGAGTCCCTACCTGATTTCTCAGGCTTGTACTCTGTGCTCAATTTGGGCCAAC 480
Db 1009 ACTGAGTCCCTACCTGATTTCTCAGGCTTGTACTCTGTGCTCAATTTGGGCCAAC 1068
Qy 481 CGAATTTCTCCCAATCTTTATCTTGCTGGCCAGGAGATGTCCTCAAGAGAGCTGATA 540
Db 1069 CGAATTTCTCCCAATCTTTATCTTGCTGGCCAGGAGATGTCCTCAAGAGAGCTGATA 1128
Qy 541 CAGCAGAAATGGGATTTGGTATGTGTAAATGCCAGCTTAACTGTCCAAAGCCCTGACTT 600
Db 1129 CAGCAGAAATGGGATTTGGTATGTGTAAATGCCAGCAATACGTGTCCAAAGCCCTGACTT 1188
Qy 601 ATCCCGAGTCTCAATTTCTGCGGTGCTGTGAATGACAGCTTTGTGAGAAAATTTTG 660
Db 1189 ATCCCGAGTCTCAATTTCTGCGGTGCTGTGAATGACAGCTTTGTGAGAAAATTTTG 1248
Qy 661 CCGTGTGTGACAAATCAGTAGATTTCATTGAGAAACAAAGCCTTCAATGGATGTGT 720
Db 1249 CCGTGTGTGACAAATCAGTAGATTTCATTGAGAAACAAAGCCTTCAATGGATGTGT 1308
Qy 721 CTAGTGACTGTTTAGTGGGATCTCCCGCTCCGCCACATGCTTATCGCTTACATCATG 780
Db 1309 CTAGTGACTGTTTAGTGGGATCTCCCGCTCCGCCACATGCTTATCGCTTACATCATG 1368
Qy 781 AAGAGATGACATGTTTGTAGTGAAGCTTACAGATTGTGAAAGAAAAGAAAGCTACT 840
Db 1369 AAGAGATGACATGTTTGTAGTGAAGCTTACAGATTGTGAAAGAAAAGAAAGCTACT 1428
Qy 841 ATATCTCCAAATCTCAATTTTCTGGGCAACCTCTGACATGAGAGAAATTTAAGAAC 900
Db 1429 ATATCTCCAAATCTCAATTTTCTGGGCAACCTCTGACATGAGAGAAATTTAAGAAC 1488
Qy 901 CAGACTGAGACATCAGGGCCAAAGAGCAAACTCAAGCTGTGACCTGAGAGAAAGCCAAAT 960
Db 1489 CAGACTGAGACATCAGGGCCAAAGAGCAAACTCAAGCTGTGACCTGAGAGAAAGCCAAAT 1548
Qy 961 GAACTGTCTCTGCTGTCTCAGAGGTGACAGAAAAGAGAGAGCCCTCACTCAACCC 1020
Db 1549 GAACTGTCTCTGCTGTCTCAGAGGTGACAGAAAAGAGAGAGCCCTCACTCAACCC 1608
Qy 1021 TGTGCGGACTGTGCTACCTCAGAGGCAAGCAAAAGCCCTGTGATCCGCCAGCGTG 1080
Db 1609 TGTGCGGACTGTGCTACCTCAGAGGCAAGCAAAAGCCCTGTGATCCGCCAGCGTG 1668
Qy 1081 CCCAGGTCGCCAGCGTGCAGCGGTGTTAGAGGACAGCCCGCTGTACAGGCGCTG 1140
Db 1669 CCCAGGTCGCCAGCGTGCAGCGGTGTTAGAGGAGAGCCCGCTGTACAGGCGCTG 1728
Qy 1141 AGTGGCTGCACCTGTGCGCAGACAGGCTGAAAGCAGCAATTAAGCTTAAAGCTTTC 1200
Db 1729 AGTGGCTGCACCTGTGCGCAGACAGGCTGAAAGCAGCAATTAAGCTTAAAGCTTTC 1788
Qy 1201 TCTCGGATATCAAAATCAGTTTCAATTCAGCCAGATGGAGCATCTCTTACATGAGCTTC 1260
Db 1789 TCTCGGATATCAAAATCAGTTTCAATTCAGCCAGATGGAGCATCTCTTACATGAGCTTC 1848
Qy 1261 TCCCTATCAGAAAGATGCTTTGGATATCTAATAAATCTTCACTACTCTGAGGGACAAC 1320
Db 1849 TCCCTATCAGAAAGATGCTTTGGATATCTAATAAATCTTCACTACTCTGAGGGACAAC 1908
Qy 1321 AAGCATATGCCAGTCTCCCTGTTTCAAGAACTATTCAGAGCAAGCTCCGAAACAGTCTCT 1380
Db 1909 AAGCATATGCCAGTCTCCCTGTTTCAAGAACTATTCAGAGCAAGCTCCGAAACAGTCTCT 1968

Qy 1381 GATTAAGAGAAACCCAGCATCTCCCAAGAAAGCTGCAGACCGCCAGGCTTACAGACGCCAG 1440
Db 1969 GATTAAGAGAAAGCCAGCATCTCCCAAGAAAGCTGCAGACCGCCAGGCTTACAGACGCCAG 2028
Qy 1441 AGCAAGCATATGATCTCCGCTCAAGAAACAGCAGCATTTGGCAGCCCAAGAGTCCCTTTTA 1500
Db 2029 AGCAAGCATATGATCTCCGCTCAAGAAACAGCAGCATTTGGCAGCCCAAGAGTCCCTTTTA 2088
Qy 1501 TCTCCAGTGCATCGAATGGGAGCGTGGAGACAAATTAACACACAGCTCTCTTTTCGAC 1560
Db 2089 TCTCCAGTGCATCGAATGGGAGCGTGGAGACAAATTAACACACAGCTCTCTTTTCGAC 2148
Qy 1561 CTTTCACAGCCAGCAGACGACTTCAAGAAATGTGTGCTGGGCTTAAAGGCTGGAC 1620
Db 2149 CTTTCACAGCCAGCAGACGACTTCAAGAAATGTGTGCTGGGCTTAAAGGCTGGAC 2208
Qy 1621 TCGGATATCTTGGCCCCCAGACCTTACCCCTTCCCTGACACGAGAGCTGATTTTGGC 1680
Db 2209 TCGGATATCTTGGCCCCCAGACCTTACCCCTTCCCTGACACGAGAGCTGATTTTGGC 2268
Qy 1681 ACAGAGTCTCAGACTTCTACTCTGCTCAAGCCTTACGAGGAGAGTCCAGTTACTCT 1740
Db 2269 ACAGAGTCTCAGACTTCTACTCTGCTCAAGCCTTACGAGGAGAGTCCAGTTACTCT 2328
Qy 1741 GCTTACAGCTCAGCAGCAGCTGCCACTTGGCGGAAACCAAGTCTTATCTGTGCGAGCGG 1800
Db 2329 GCTTACAGCTCAGCAGCAGCTGCCACTTGGCGGAAACCAAGTCTTATCTGTGCGAGCGG 2388
Qy 1801 CAGAAAGCAATGACAGAGCTGACTGCGCGGAGAGCTGAGCAATGAGAGAGGCGCTTTGAA 1860
Db 2389 CAGAAAGCAATGACAGAGCTGACTGCGCGGAGAGCTGAGCAATGAGAGAGGCGCTTTGAA 2448
Qy 1861 AAGCATTTAAACCGAGAACTGCCAAATGGAATTTGAGAGAGCATCATGTCAAGAAC 1920
Db 2449 AAGCATTTAAACCGAGAACTGCCAAATGGAATTTGAGAGAGCATCATGTCAAGAAC 2508
Qy 1921 AGTTCAGGGAAGAGCTGGGGAAGTGGGAGTCAAGTCAAGTCTTTGGGCAAGATGGA 1980
Db 2509 AGTTCAGGGAAGAGCTGGGGAAGTGGGAGTCAAGTCTTTGGGCAAGATGGA 2568
Qy 1981 ATCATTTGAGTCTCC 1995
Db 2569 ATCATTTGAGTCTCC 2583

RESULT 18
ABV26923
ID ABV26923 standard; cDNA; 5145 BP.
XX
AC ABV26923;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 26914.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX
OS pharmacogenomic marker; gene; ss.
XX
Homo sapiens.
XX
PN W0200160860-A2.
PD
XX
23-AUG-2001.
PF
20-FEB-2001; 2001MO-US005171.
XX
17-FEB-2000; 2000US-0183119P.
PR
16-MAR-2000; 2000US-0189862P.
PR
25-MAY-2000; 2000US-0207544P.
PR
09-JUN-2000; 2000US-0211314P.
PR
18-JUL-2000; 2000US-0219007P.
PR
13-DEC-2000; 2000US-0255281P.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PA Schlegel R, Endege WO, Monahan JR;
PI MPI; 2001-662795/76.
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
XX Claim 1; Page 5449-5450; 11750pp; English.
XX The invention relates to an isolated nucleic acid molecule (1) comprising
CC a nucleotide sequence given in Tables 1-9 (ABY0010-ABY62213) of the
CC specification or its complement. (1) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
XX Sequence 5145 BP; 1408 A; 1135 C; 1253 G; 1346 T; 0 U; 3 Other;
SQ
Query Match 94.9%; Score 1893; DB 5; Length 5145;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGGCCCATGATGATGGAATCAATGTTTACTGAGAGTTGGTGGCTCTGCTGAA 60
DB 589 ATGGCCCATGATGATGGAATCAATGTTTACTGAGAGTTGGTGGCTCTGCTGAA 648
QY 61 AGTGAACCGAAGAGGCTGCTAATTGATAGCCGCAATTGTGGAATCATATAC 120
DB 649 AGTGAACCGAAGAGGCTGCTAATTGATAGCCGCAATTGTGGAATCATATAC 708
QY 121 CACATTTTGGAGCCATTAATATCACTGCTCAAGCTTATGAAGGAGTTGCAAC 180
DB 709 CACATTTTGGAGCCATTAATATCACTGCTCAAGCTTATGAAGGAGTTGCAAC 768
QY 181 GACAAAGTTAATACAGAGCTCATCCAGCATTCAGCGAACAATAAGTTGACATT 240
DB 769 GACAAAGTTAATACAGAGCTCATCCAGCATTCAGCGAACAATAAGTTGACATT 828
QY 241 TGCAGTCAGAAAGTTGATTAAGATCAAAAGCTCCCAAGATGTTGCTCTCTTCA 300
DB 829 TGCAGTCAGAAAGTTGATTAAGATCAAAAGCTCCCAAGATGTTGCTCTCTTCA 888
QY 301 GACTGTTTCTCACTGACTTCTGGGTAATCTGAGAAAGAGTTCAACTCTGTTCA 360
DB 889 GACTGTTTCTCACTGACTTCTGGGTAATCTGAGAAAGAGTTCAACTCTGTTCA 948
QY 361 CTTCAGAGTGGGTTTGTGAGTTCTCTGTTGTTTCCCTGAGCTCTGTGAGAAAT 420
DB 949 CTTCAGAGTGGGTTTGTGAGTTCTCTGTTGTTTCCCTGAGCTCTGTGAGAAAT 1008
QY 421 ACTCTAGTCCCTTACCTGATTTCTCAGCCTTGTACCTGTGACCAATTTGGCC 480
DB 1009 ACTCTAGTCCCTTACCTGATTTCTCAGCCTTGTACCTGTGACCAATTTGGCC 1068
QY 481 CGAATTTTCCCAATCTTTATCTTGGCTGCAGAGAGATGCTCTCAACAGAGCTG 540
DB 1069 CGAATTTTCCCAATCTTTATCTTGGCTGCAGAGAGATGCTCTCAACAGAGCTG 1128
QY 541 CAGCAGATGAGATGTTATGTTTAAATGCAAGCTATACCTGTCCAAAGCTGACT 600
DB 1129 CAGCAGATGAGATGTTATGTTTAAATGCAAGCTATACCTGTCCAAAGCTGACT 1188
QY 601 ATCCCGAAGTCTATTCTCGCGTGTGCTGTGAATACAGCTTTTGTGAGAAAT 660

DB 1189 ATCCCGAAGTCTATTCTCGGTGTGCTGTGAATGACAGCTTTTGTGAGAAAT 1248
QY 661 CCGTGTGGACCAATCAGTATGATTTTATTGGAAGCAAAAGCTCTCAATGATGTT 720
DB 1249 CCGTGTGGACCAATCAGTATGATTTTATTGGAAGCAAAAGCTCTCAATGATGTT 1308
QY 721 CTAGTGCATGTTTATGCTGGGATCTCCGCTCCGACCATGATGCTTACATCAT 780
DB 1309 CTAGTGCATGTTTATGCTGGGATCTCCGCTCCGACCATGATGCTTACATCAT 1368
QY 781 AAGAGATGAGATGCTTTTATGATGAAGCTTACAGATTTGTGAGAAAGAAAGCT 840
DB 1369 AAGAGATGAGATGCTTTTATGATGAAGCTTACAGATTTGTGAGAAAGAAAGCT 1428
QY 841 ATATCTCCAAATTTTCTGGGCACTCTGAGTATGAGAAAGATTAAGAAC 900
DB 1429 ATATCTCCAAATTTTCTGGGCACTCTGAGTATGAGAAAGATTAAGAAC 1488
QY 901 CAGCTGAGCATCAGGCGCAAGCAAACTCAAGCTGCTGACCTGAGAAAGCCAAAT 960
DB 1489 CAGCTGAGCATCAGGCGCAAGCAAACTCAAGCTGCTGACCTGAGAAAGCCAAAT 1548
QY 961 GAACCTGCTCTGCTGCTCTCAGAGGAGTGAAGAAAGAGAGAGAGAGAGAGAG 1020
DB 1549 GAACCTGCTCTGCTGCTCTCAGAGGAGTGAAGAAAGAGAGAGAGAGAGAGAG 1608
QY 1021 TGTGCGCACTGCTACTCCTCAGAGGAGAGAGCAAAAGCCGTGATCCGCAAGCT 1080
DB 1609 TGTGCGCACTGCTACTCCTCAGAGGAGAGAGCAAAAGCCGTGATCCGCAAGCT 1668
QY 1081 CCCAGGTCGCCAGGCTGAGAGCCGCTGTTAGAGAGAGAGAGAGAGAGAGAGAG 1140
DB 1669 CCCAGGTCGCCAGGCTGAGAGCCGCTGTTAGAGAGAGAGAGAGAGAGAGAGAG 1728
QY 1141 AGTGGCTGCACTGCTGCTCAGAGGAGTGAAGAGAGAGAGAGAGAGAGAGAG 1200
DB 1729 AGTGGCTGCACTGCTGCTCAGAGGAGTGAAGAGAGAGAGAGAGAGAGAGAG 1788
QY 1201 TCTCTGATATCAATCAATCAATTTATATTCAGCCAGATGAGCATCTTACATG 1260
DB 1789 TCTCTGATATCAATCAATCAATTTATATTCAGCCAGATGAGCATCTTACATG 1848
QY 1261 TCTCTATCAAGAGTCTTTGGAATCTAAGAACTTCCATCTCTGATGAGAGAG 1320
DB 1849 TCTCTATCAAGAGTCTTTGGAATCTAAGAACTTCCATCTCTGATGAGAGAG 1908
QY 1321 AAGCTATGCAAGTCTCCCTGCTTCAAGAACTATGAGAGAGAGAGAGAGAGAG 1380
DB 1909 AAGCTATGCAAGTCTCCCTGCTTCAAGAACTATGAGAGAGAGAGAGAGAGAG 1968
QY 1381 GATTAAGAGAGAGAGAGAGATCCCAAGAGCTGAGAGAGAGAGAGAGAGAGAG 1440
DB 1969 GATTAAGAGAGAGAGAGATCCCAAGAGCTGAGAGAGAGAGAGAGAGAGAGAG 2028
QY 1441 AGCAAGAGATGATGCTTGGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
DB 2029 AGCAAGAGATGATGCTTGGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2088
QY 1501 TCTCCAGTCAAG 1560
DB 2089 TCTCCAGTCAAG 2148
QY 1561 CTTCACAG 1620
DB 2149 CTTCACAG 2208
QY 1621 TGGAGATCTTGGAG 1680
DB 2209 TGGAGATCTTGGAG 2268
QY 1681 ACAGAGTCTCACTTACTCTGAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1740

CC	assessing the aggressiveness or incidence of prostate cancer in a patient
CC	; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
XX	
XX	
SQ	Sequence 5145 BP; 1408 A; 1135 C; 1253 G; 1346 T; 0 U; 3 Other;
	Query Match
	Best Local Similarity 94.9%; Score 1893; DB 5; Length 5145;
	Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0
OY	1 ATGGCCCATGATGATATTGGAACCAATTTGTACTAGAGTGTGGCGTCGTCTGGAA 60
Db	589 ATGGCCCATGAGATATTTGGAACTCCAATTTTACTAGAGTTTGCTGCTCTGGAA 648
OY	61 AGTGGAAACGAAAAAGTCTCTTAATTGATAGCCGGCATTITGTGNAATACAATCATCC 120
Db	649 AGTGGAAACGAAAAAGTCTCTTAATTGATAGCCGGCATTITGTGNAATACAATCATCC 708
OY	121 CACATTTTGGAAAGCATTAAATATAACTGCTCCAAAGTTATGAAGGAAGTTTGCACAG 180
Db	709 CACATTTTGGAAAGCATTAAATATAACTGCTCCAAAGTTATGAAGGAAGTTTGCACAG 768

1

2

?

25

300

02

Q

100
 nh

4807

106
Ph

0V 54

Dib 112

0v 60

Db 118

66
Oy

Db 124

72 QY

Db 130

QY 78

Db 136

84 QY

Db 1421

90: QY

1489 Db

QY 961 GAACCTGCTCCCTGCTCTCTCAGAGGGTGAACAGAAAAAGCAGAGCCCTCACTCCACCC 1020
DB 1549 GAACTGTGCTCTGCTCTCTCAGAGGGTGAACAGAAAAAGCAGAGCCCTCACTCCACCC 1608
QY 1021 TGTGCGGACTCTGCTCTCTCAGAGGGTGAACAGAAAAAGCAGAGCCCTCACTCCACCC 1080
DB 1609 TGTGCGGACTCTGCTCTCTCAGAGGGTGAACAGAAAAAGCAGAGCCCTCACTCCACCC 1668
QY 1081 CCCAGGCTGCGCAGCGGTCAGCGGCTGCTGAGAGGACAGCGCGCTGAGAGCGGCTC 1140
DB 1669 CCCAGGCTGCGCAGCGGTCAGCGGCTGCTGAGAGGACAGCGCGCTGAGAGCGGCTC 1728
QY 1141 AGTGGGCTGCACTGCTCTCAGAGGGTGAACAGAAAAAGCAGAGCCCTCACTCCACCC 1200
DB 1729 AGTGGGCTGCACTGCTCTCAGAGGGTGAACAGAAAAAGCAGAGCCCTCACTCCACCC 1788
QY 1201 TCTCTGAGATCAAAATCAGTTTCAATTCAGCCAGAGGAGCATCTTCAATGAGCTTC 1260
DB 1789 TCTCTGAGATCAAAATCAGTTTCAATTCAGCCAGAGGAGCATCTTCAATGAGCTTC 1848
QY 1261 TCTCTCAACAGAAATGCTTGGAAATCTAATAAATCTTCCATCTCTGAGTGGGACCAAC 1320
DB 1849 TCTCTCAACAGAAATGCTTGGAAATCTAATAAATCTTCCATCTCTGAGTGGGACCAAC 1908
QY 1321 AAGCTATGCGAGTCTCTCCCTGTTCCAGAACTATCGGAGAGCATCCCGAAACAGTCTCT 1380
DB 1909 AAGCTATGCGAGTCTCTCCCTGTTCCAGAACTATCGGAGAGCATCCCGAAACAGTCTCT 1968
QY 1381 GATTAAGAGAGAGCCAGCATCCCAAGAACTGTCAGAGCGGCTTTCAGACAGCCAG 1440
DB 1969 GATTAAGAGAGAGCCAGCATCCCAAGAACTGTCAGAGCGGCTTTCAGACAGCCAG 2028
QY 1441 AGCAAGCATGATGATTCGGTCAAGAACAGAGAGAGGAGCCCGCCAGAGAGTCCCTTTTA 1500
DB 2029 AGCAAGCATGATGATTCGGTCAAGAACAGAGAGAGGAGCCCGCCAGAGAGTCCCTTTTA 2088
QY 1501 TCTCCACTGATGAGAGTGGAGAGCGTGGAGAGCAATTAACAACAGCTTCTTTTCGCG 1560
DB 2089 TCTCCACTGATGAGAGTGGAGAGCGTGGAGAGCAATTAACAACAGCTTCTTTTCGCG 2148
QY 1561 CTTTTCACAGCAGCAGCAGCAGCCTCACGAACTGCTGCGCTGGGCTTTAAAGGCTGGCAG 1620
DB 2149 CTTTTCACAGCAGCAGCAGCAGCCTCACGAACTGCTGCGCTGGGCTTTAAAGGCTGGCAG 2208
QY 1621 TCGGATATCTTGGCCCCCAGAGACTCTAATCCCTTCCCTGACAGAGCTGGATTTTGGC 1680
DB 2209 TCGGATATCTTGGCCCCCAGAGACTCTAATCCCTTCCCTGACAGAGCTGGATTTTGGC 2268
QY 1681 ACAGAGTCTCACTCTCTCTGCTGCTGAGCATCTAAGGAGAGGAGTGGCAGTTAATCTCT 1740
DB 2269 ACAGAGTCTCACTCTCTCTGCTGCTGAGCATCTAAGGAGAGGAGTGGCAGTTAATCTCT 2328
QY 1741 GCCTAAGCTGACAGCAGCTGCCACTTGGCGAGAGCAATCTAATCTGTCGACAGGCGG 1800
DB 2329 GCCTAAGCTGACAGCAGCTGCCACTTGGCGAGAGCAATCTAATCTGTCGACAGGCGG 2388
QY 1801 CAGAGCCAGAGTGAAGAGCTGAGCTGCGCGGAGAGCTGCGATGAAGAGAGCCCTTTGAA 1860
DB 2389 CAGAGCCAGAGTGAAGAGCTGAGCTGCGCGGAGAGCTGCGATGAAGAGAGCCCTTTGAA 2448
QY 1861 AAGCAGTTTAAAGCAGAGCTGCCAAATGGAATTTGAGAGAGCATATGTCAGAGAAC 1920
DB 2449 AAGCAGTTTAAAGCAGAGCTGCCAAATGGAATTTGAGAGAGCATATGTCAGAGAAC 2508
QY 1921 AGGTCAAGGAGAGAGCTGGGAGAAAGTGGGAGTCAAGTCTAAGCTTTTCGCGCAGCATGGA 1980
DB 2509 AGGTCAAGGAGAGAGCTGGGAGAAAGTGGGAGTCAAGTCTAAGCTTTTCGCGCAGCATGGA 2568
QY 1981 ATCATTGAGGTCTCC 1995
DB 2569 ATCATTGAGGTCTCC 2583

RESULT 20
ABV28657
ID ABV28657 standard; cDNA; 5145 BP.
XX
AC ABV28657;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker CDNA 28648.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN W0200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001M0-US005171.
XX
PR 17-FEB-2000; 2000US-018319P.
XX
PR 16-MAR-2000; 2000US-0189862P.
XX
PR 25-MAY-2000; 2000US-0207454P.
XX
PR 09-JUN-2000; 2000US-0211314P.
XX
PR 18-JUL-2000; 2000US-0219007P.
XX
PR 13-DEC-2000; 2000US-0255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS Claim 1, Page 6009-6010, 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (i) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (i) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
SQ Sequence 5145 BP; 1408 A; 1135 C; 1253 G; 1346 T; 0 U; 3 Other;
XX
Query Match 94.9%; Score 1893; DB 5; Length 5145;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGGCCCATGAGATGATTTGGAAGTCAAAATTTGTAAGAGAGGTTGGCTGCTGGA 60
DB 589 ATGGCCCATGAGATGATTTGGAAGTCAAAATTTGTAAGAGAGGTTGGCTGCTGGA 648
QY 61 AGTGAACGGAAGAAAGTGTCTTAATTAATGATCCGCGCATTTTGGAAATACATATCC 120
DB 649 AGTGAACGGAAGAAAGTGTCTTAATTAATGATCCGCGCATTTTGGAAATACATATCC 708
QY 121 CACATTTTGAAGGCAATTAATTAATGATCCGCGCATTTTGAAGAGGTTGAACAG 180
DB 709 CACATTTTGAAGGCAATTAATTAATGATCCGCGCATTTTGAAGAGGTTGAACAG 768
QY 181 GACAAAGTGTATTAATTAAGAGCTCAATCCAGATTCAGCGAAATTAAGTTGACATTCAT 240

Db 769 GACAAAGTGTAAATTACAGAGCTCATCCAGCATTCAGGCAAAACATAAGATTGACATTGAT 828
Qy 241 TGCAGTCAGAGAGTGTGATTAGCATGCAAAAGCTCCCAAGATGTTGCTCTCTCTTCA 300
Db 829 TGCAGTCAGAGAGTGTGATTAGCATGCAAAAGCTCCCAAGATGTTGCTCTCTCTTCA 888
Qy 301 GACTGTTTTCTGACTGTAATTCTGGTAAACTGAGAGAGCTTCAACTCTGTTCACTG 360
Db 889 GACTGTTTTCTGACTGTAATTCTGGTAAACTGAGAGAGCTTCAACTCTGTTCACTG 948
Qy 361 CTGAGAGTGGTGTGCTGAGTCTCTGTTGTTTCCCTGGCTCTGTGAAGAAAAATCC 420
Db 949 CTGAGAGTGGTGTGCTGAGTCTCTGTTGTTTCCCTGGCTCTGTGAAGAAAAATCC 1008
Qy 421 ACTTAGTCCCTACCTGCAATTTCTCAGGCTTGAACCTGTTGCAACATTGGGCCAAC 480
Db 1009 ACTTAGTCCCTACCTGCAATTTCTCAGGCTTGAACCTGTTGCAACATTGGGCCAAC 1068
Qy 481 CGAATTTCTCCCAATCTTTATCTTGGCTGCCAGAGATGTCCTCAACAGAGCTGATA 540
Db 1069 CGAATTTCTCCCAATCTTTATCTTGGCTGCCAGAGATGTCCTCAACAGAGCTGATA 1128
Qy 541 CAGCAGATGGGATTTGTTATGTTAAATGCGCAGCTATACCTGTCCAAAGCTGACTTT 600
Db 1129 CAGCAGATGGGATTTGTTATGTTAAATGCGCAGCAATACCTGTCCAAAGCTGACTTT 1188
Qy 601 ATCCCGAGTCTATTTCTGCGTGTGCTGCTGATGACAGCTTTTGTGGAATAATTTG 660
Db 1189 ATCCCGAGTCTATTTCTGCGTGTGCTGCTGATGACAGCTTTTGTGGAATAATTTG 1248
Qy 661 CCGTGTGTGA CAATCATGTAATTTATGTAAGAAAGCAAAAGCTTCAATGGATGTT 720
Db 1249 CCGTGTGTGA CAATCATGTAATTTATGTAAGAAAGCAAAAGCTTCAATGGATGTT 1308
Qy 721 CTAGTGACTGTTAGTGGGATCTCCGCTCCGCCACCATGGCTATGCCCTACATCAG 780
Db 1309 CTAGTGACTGTTAGTGGGATCTCCGCTCCGCCACCATGGCTATGCCCTACATCAG 1368
Qy 781 AAGAGATGACATGTCCTTTAGATGAAGCTTACAGATTTGTGAAAGAAAAAGACTACT 840
Db 1369 AAGAGATGACATGTCCTTTAGATGAAGCTTACAGATTTGTGAAAGAAAAAGACTACT 1428
Qy 841 ATATCTCCAAACTCAATTTCTGGGCCAACTCTCTGACTATGAGAAAGATTAAAGAC 900
Db 1429 ATATCTCCAAACTCAATTTCTGGGCCAACTCTCTGACTATGAGAAAGATTAAAGAC 1488
Qy 901 CAGACTGAGCATCAGGGCCAAAGAGCAAACTCAGCTGCTGCACTGGAGAGGCAAT 960
Db 1489 CAGACTGAGCATCAGGGCCAAAGAGCAAACTCAGCTGCTGCACTGGAGAGGCAAT 1548
Qy 961 GAACTGTCTCTGCTCTCTCAAGGGTGAAGAAAGGAGAGCGCCCTCAGTCAACC 1020
Db 1549 GAACTGTCTCTGCTCTCTCAAGGGTGAAGAAAGGAGAGCGCCCTCAGTCAACC 1608
Qy 1021 TGTGCGACTGTGCTCACTCAGAGGAGAGGCAAAAGGCGCTGCACTCCGCGAGCTG 1080
Db 1609 TGTGCGACTGTGCTCACTCAGAGGAGAGGCAAAAGGCGCTGCACTCCGCGAGCTG 1668
Qy 1081 CCCAGGTGCGCAGCGGTGAGCGCTGTGTTAGAGAGCAGCCGCTGTTACAGGCGCTC 1140
Db 1669 CCCAGGTGCGCAGCGGTGAGCGCTGTGTTAGAGAGCAGCCGCTGTTACAGGCGCTC 1728
Qy 1141 AGTGGGCTGACCTGTCCGAGAGAGGCTGAGAGCAGCAATAGCTCAGCGTTCTTC 1200
Db 1729 AGTGGGCTGACCTGTCCGAGAGAGGCTGAGAGCAGCAATAGCTCAGCGTTCTTC 1788
Qy 1201 TCTGTGATATCAATCAATGTTTCAATTCAGCAGCATGGAGATCCTTACAGGCTTC 1260
Db 1789 TCTGTGATATCAATCAATGTTTCAATTCAGCAGCATGGAGATCCTTACAGGCTTC 1848
Qy 1261 TCCATATCAAGAGATGCTTTGGAATCTACAAACCTTTCAGTACTGTGATGGAGCAAC 1320
Db 1849 TCCATATCAAGAGATGCTTTGGAATCTACAAACCTTTCAGTACTGTGATGGAGCAAC 1908

Qy 1321 AAGCTATGCAAGTTCCTCCCTGTTGAGGAATATGCAAGACATCCGGAACAGTCT 1380
Db 1909 AAGCTATGCAAGTTCCTCCCTGTTGAGGAATATGCAAGACATCCGGAACAGTCT 1968
Qy 1381 GATTAAGAGAGAGCCAGCATCTCCCAAGAAAGCTGACAGCCGCTTACAGACAGCAG 1440
Db 1969 GATTAAGAGAGAGCCAGCATCTCCCAAGAAAGCTGACAGCCGCTTACAGACAGCAG 2028
Qy 1441 ACCAAGCATTCATTCGGTCAAGACAGAGCGATGGCAACCGCCAGAGTCCCTTTTA 1500
Db 2029 ACCAAGCATTCATTCGGTCAAGACAGAGCGATGGCAACCGCCAGAGTCCCTTTTA 2088
Qy 1501 TCTCCACTGCAATCGAAGTGGAGCGGTGAGAGCAATTAACAACAGCTTCTTTGGGC 1560
Db 2089 TCTCCACTGCAATCGAAGTGGAGCGGTGAGAGCAATTAACAACAGCTTCTTTGGGC 2148
Qy 1561 CTTTCACACAGCCAGCAGCAGCAGCTTCAACGAAGTCTGCTGGCTTTAAGGGCTGGCAC 1620
Db 2149 CTTTCACACAGCCAGCAGCAGCAGCTTCAACGAAGTCTGCTGGCTTTAAGGGCTGGCAC 2208
Qy 1621 TCGATATCTTGGCCGCCAGACCTTACCCCTTCCGACAGCAGCTGGTATTTTGGC 1680
Db 2209 TCGATATCTTGGCCGCCAGACCTTACCCCTTCCGACAGCAGCTGGTATTTTGGC 2268
Qy 1681 ACAGAGTCTTCAACTTCTACTGCTGCTCAGCACTTAACGAGGCACTGCACTTACT 1740
Db 2269 ACAGAGTCTTCAACTTCTACTGCTGCTCAGCACTTAACGAGGCACTGCACTTACT 2328
Qy 1741 GCTTACAGCTGACAGCTGCGCCACTTGGGGAGACCAATCTATTTCTGTGCGCAGGCG 1800
Db 2329 GCTTACAGCTGACAGCTGCGCCACTTGGGGAGACCAATCTATTTCTGTGCGCAGGCG 2388
Qy 1801 CAGAACCAAGTGA CAGAGCTGA CTGCGCGGAGCTGCAATGAAGAGAGCCCTTTGAA 1860
Db 2389 CAGAACCAAGTGA CAGAGCTGA CTGCGCGGAGCTGCAATGAAGAGAGCCCTTTGAA 2448
Qy 1861 AAGCATTTTAAACGCAAACTGCTCAATGGAATTTGGAGAGCATATGTCAGAGAAC 1920
Db 2449 AAGCATTTTAAACGCAAACTGCTCAATGGAATTTGGAGAGCATATGTCAGAGAAC 2508
Qy 1921 AGGTCAAGGAGAGCTGGGAGAAAGTGGGAGTCACTAGCTTTTGGGCGAGCATGGAA 1980
Db 2509 AGGTCAAGGAGAGCTGGGAGAAAGTGGGAGTCACTAGCTTTTGGGCGAGCATGGAA 2568
Qy 1981 ATCATTGAGTCTCC 1995
Db 2569 ATCATTGAGTCTCC 2583

RESULT 21
ABV22827 standard; cDNA, 5145 BP.
XX
AC ABV22827;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cdna 22818.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN MO200160860-A2.
PD
XX
PD 23-AUG-2001.
PF
XX
PF 20-FEB-2001; 2001WO-US005171.
PR
XX
PR 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.

Db 2209 TCGATATCTTGGCCCCCAGACCTCTACCCCTTCCCTGACAGAGCTGATTTTGGC 2268
Qy 1661 ACAGAGTCTCACTTCTACTCTGCTCCCTACCCATTTACGAGGAGTGGCCATTTCTT 1740
Db 2269 ACAGAGTCTCACTTCTACTCTGCTCCCTACCCATTTACGAGGAGTGGCCATTTCTT 2328
Qy 1741 GCCTACAGCTCAGGACCTGCTCCCACTTGGCGAAGCAAGTCTATTCTGTGGCGAGGCGG 1800
Db 2329 GCCTACAGCTCAGGACCTGCTCCCACTTGGCGAAGCAAGTCTATTCTGTGGCGAGGCGG 2388
Qy 1801 CAGAGCCCAATGACAGAGCTGCTCCGCGCGGAGCTGGCAATGAGAGAGCCCTTTGAA 1860
Db 2389 CAGAGCCCAATGACAGAGCTGCTCCGCGCGGAGCTGGCAATGAGAGAGCCCTTTGAA 2448
Qy 1861 AAGCAGTTTAAACCCAGAGAGCTGCGCAATGGAATTTGAGAGAGCATGATGAGAGAC 1920
Db 2449 AAGCAGTTTAAACCCAGAGAGCTGCGCAATGGAATTTGAGAGAGCATGATGAGAGAC 2508
Qy 1921 AGGTCAAGGAGAGAGCTGGGGAAGTGGGAGTCACTAGCTTTTGGGCGAGCATGGA 1980
Db 2509 AGGTCAAGGAGAGAGCTGGGGAAGTGGGAGTCACTAGCTTTTGGGCGAGCATGGA 2568
Qy 1981 ATCATTAGGTCTCC 1995
Db 2569 ATCATTAGGTCTCC 2583

RESULT 22

ABV26934
ID ABV26934 standard; cDNA; 5145 BP.

XX ABV26934;

DT 16-SEP-2002 (first entry)

DE Human prostate expression marker cDNA 26925.

KM Human; prostate cancer; cytosolic; carcinogen; pharmacodynamic marker;

KW pharmacogenomic marker; gene; ss.

OS Homo sapiens.

PN W0200160860-A2.

PD 23-AUG-2001.

PF 20-FEB-2001; 2001WO-US005171.

PR 17-FEB-2000; 2000US-0183319P.

PR 16-MAR-2000; 2000US-0189862P.

PR 25-MAY-2000; 2000US-0207454P.

PR 09-JUN-2000; 2000US-0211314P.

PR 18-JUL-2000; 2000US-0219007P.

PR 13-DEC-2000; 2000US-0255281P.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

PI Schlegel R, Endege WO, Monahan JE;

XX WPI; 2001-662795/76.

XX Novel isolated nucleic acid molecule associated with cancerous state of

PT prostate cells and correlating with presence of prostate cancer, useful

CC for detecting presence of prostate cancer, stage of prostate cancer.

CC Claim 1; Page 5453-5454; 11750pp; English.

CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker
XX

SQ Sequence 5145 BP; 1408 A; 1135 C; 1253 G; 1346 T; 0 U; 3 Other;

Query Match 94.9%; Score 1893; DB 5; Length 5145;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGCCCATGAGATATTGGAATCTCAATTTGTAATGAGAGTTGGTCTCTGTGGA 60
Db 589 ATGGCCCATGAGATATTGGAATCTCAATTTGTAATGAGAGTTGGTCTCTGTGGA 648
Qy 61 AGTGAACGGAAAAAGTCTGCTAATTGATGACCGGCGCATTTTGTGGAATACATATCC 120
Db 649 AGTGAACGGAAAAAGTCTGCTAATTGATGACCGGCGCATTTTGTGGAATACATATCC 708
Qy 121 CACATTTTGAAGGACATTAATATCACTGCTCCAAAGCTTATGAGCGAAGGTTGCAACAG 180
Db 709 CACATTTTGAAGGACATTAATATCACTGCTCCAAAGCTTATGAGCGAAGGTTGCAACAG 768
Qy 181 GACAAAGTGTAAATTACAGAGCTCATTCAGCATTCAGCGAAACATAAGTTGACATTGAT 240
Db 769 GACAAAGTGTAAATTACAGAGCTCATTCAGCATTCAGCGAAACATAAGTTGACATTGAT 828
Qy 241 TGCAGTGAAGAGTTGATGATTTAGATCAAAAGCTCCCAAGATGTTGGCTCTCTCTTCA 300
Db 829 TGCAGTGAAGAGTTGATGATTTAGATCAAAAGCTCCCAAGATGTTGGCTCTCTCTTCA 888
Qy 301 GACTGTTTCTCAGCTGATCTTGGGTAACTGGAAGAGAGCTTCACTGTTCACCTG 360
Db 889 GACTGTTTCTCAGCTGATCTTGGGTAACTGGAAGAGAGCTTCACTGTTCACCTG 948
Qy 361 CTTCAGAGTGGGTTGCTGAGTTCTCTGTTGTTTCCCTGGCTCTGTGAAGAAATCC 420
Db 949 CTTCAGAGTGGGTTGCTGAGTTCTCTGTTGTTTCCCTGGCTCTGTGAAGAAATCC 1008
Qy 421 ACTCTAGTCCCTACCTGATCTTCTCAGCCTTGTGCTTACCTGTGCAACATTTGGGCAACC 480
Db 1009 ACTCTAGTCCCTACCTGATCTTCTCAGCCTTGTGCTTACCTGTGCAACATTTGGGCAACC 1068
Qy 481 GGAATTTCTCCCAATCTTATCTTGGCTGCGACGAGATGCTCTCAACAAGAGCTGATA 540
Db 1069 GGAATTTCTCCCAATCTTATCTTGGCTGCGACGAGATGCTCTCAACAAGAGCTGATA 1128
Qy 541 CAGCAGAAATGGGATTTGTTATGTTAAATGCACTATACCTGTCCAAAGCTGACTTT 600
Db 1129 CAGCAGAAATGGGATTTGTTATGTTAAATGCACTATACCTGTCCAAAGCTGACTTT 1188
Qy 601 ATCCCGAGTCTGATTTCTGCTGCTGCTGCTGATGACAGCTTTGTGAGAAATTTTG 660
Db 1189 ATCCCGAGTCTGATTTCTGCTGCTGCTGCTGATGACAGCTTTGTGAGAAATTTTG 1248
Qy 661 CCGTGGTTGACAAATGATGATTTCAATTGAGAAAGCAAAAGCTCCAAATGATGTGT 720
Db 1249 CCGTGGTTGACAAATGATGATTTCAATTGAGAAAGCAAAAGCTCCAAATGATGTGT 1308
Qy 721 CTAGTCACTGTTTACTGAGGATCTCCGCTCCGCAACATGCTATGCTTCAATCATG 780
Db 1309 CTAGTCACTGTTTACTGAGGATCTCCGCTCCGCAACATGCTATGCTTCAATCATG 1368
Qy 781 AAGAGATGACATGCTTTAGATGAGCTTACAGATTTGTGAAGAAAAAGACCTACT 840
Db 1369 AAGAGATGACATGCTTTAGATGAGCTTACAGATTTGTGAAGAAAAAGACCTACT 1428
Qy 841 ATATCTCCAAATCTTCAATTTTGGGCGCAACTCTGAGCTATGAGAGAAATTAAGAAC 900
Db 1429 ATATCTCCAAATCTTCAATTTTGGGCGCAACTCTGAGCTATGAGAGAAATTAAGAAC 1488

QY 901 CAGACTGAGCATCAGGCGCCAAAGAGCAAACTCAGCTGCTGACCTGGAGAGCAAAAT 960
DB 1489 CAGACTGAGCATCAGGCGCCAAAGAGCAAACTCAGCTGCTGACCTGGAGAGCAAAAT 1548
QY 961 GAACTGTCCCTGCTGTCTCAGAGGGTGGACAGAAAAAGCAGAGCGCCCTCACTCCACCC 1020
DB 1549 GAACTGTCCCTGCTGTCTCAGAGGGTGGACAGAAAAAGCAGAGCGCCCTCACTCCACCC 1608
QY 1021 TGTGCGGACTGTCTGTCTCAGAGGACAGAGCAAAAGCCCTGCACTCCCGCAGCGTG 1080
DB 1609 TGTGCGGACTGTCTGTCTCAGAGGACAGAGCAAAAGCCCTGCACTCCCGCAGCGTG 1668
QY 1081 CCGAGGGTGGCCGAGCGGTGACGCGGTGTGAGAGGACGCGCGGTGACAGGAGGCTC 1140
DB 1669 CCGAGGGTGGCCGAGCGGTGACGCGGTGTGAGAGGACGCGCGGTGACAGGAGGCTC 1728
QY 1141 AGTGGGCTGACCTGTCCGAGACAGGCTGGAAGAGCAGCAATTAAGCTCAAGCGTCTTC 1200
DB 1729 AGTGGGCTGACCTGTCCGAGACAGGCTGGAAGAGCAGCAATTAAGCTCAAGCGTCTTC 1788
QY 1201 TCTCTGAGATTCAAATCAGTTTCAATTCACGCCAGATGCGACATCTTCACTGCTTC 1260
DB 1789 TCTCTGAGATTCAAATCAGTTTCAATTCACGCCAGATGCGACATCTTCACTGCTTC 1848
QY 1261 TCTCTCAACGAAATGCTTGGAAATCTAACAACCTTCCACTCTCTGATGGGACCAAC 1320
DB 1849 TCTCTCAACGAAATGCTTGGAAATCTAACAACCTTCCACTCTCTGATGGGACCAAC 1908
QY 1321 AAGCTATGCGAGTCTCCCTGTTCAAGAACTATCGAGAGAGTCCCGAAACAGTCTCT 1380
DB 1909 AAGCTATGCGAGTCTCCCTGTTCAAGAACTATCGAGAGAGTCCCGAAACAGTCTCT 1968
QY 1381 GATTAAGAGAGAACCGACATCCCAAGAAAGCTGACAGCCGACGCTTTCAGACCCAG 1440
DB 1969 GATTAAGAGAGAACCGACATCCCAAGAAAGCTGACAGCCGACGCTTTCAGACCCAG 2028
QY 1441 AGAAACGATTTGATTCGGTCAAGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
DB 2029 AGAAACGATTTGATTCGGTCAAGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2088
QY 1501 TCTCCACTGATCGAAGTGGGAGCGTGGAGAGCAATTAACACAGAGCTTCTTTTCGCG 1560
DB 2089 TCTCCACTGATCGAAGTGGGAGCGTGGAGAGCAATTAACACAGAGCTTCTTTTCGCG 2148
QY 1561 CTTTCCACGAGCCAGACAGCACTCAAGAACTGTGCGCTGCGCTTAAAGGGCTGCGAC 1620
DB 2149 CTTTCCACGAGCCAGACAGCACTCAAGAACTGTGCGCTGCGCTTAAAGGGCTGCGAC 2208
QY 1621 TCGGATATCTTGGCCCCCAGACCTCTACCCCTTCCCTGACAGAGCTGTGTAATTTTGGC 1680
DB 2209 TCGGATATCTTGGCCCCCAGACCTCTACCCCTTCCCTGACAGAGCTGTGTAATTTTGGC 2268
QY 1681 ACAGAGCTCTCAGACTTCTACTGTGCTCAGCCATCTACGAGGAGCGAGTGCAGTCTCT 1740
DB 2269 ACAGAGCTCTCAGACTTCTACTGTGCTCAGCCATCTACGAGGAGCGAGTGCAGTCTCT 2328
QY 1741 GCTCAAGCTGACAGCCAGCTGCCACTTGGCGAGACCAATCTATTCTGTGCGCAGAGCGG 1800
DB 2329 GCTCAAGCTGACAGCCAGCTGCCACTTGGCGAGACCAATCTATTCTGTGCGCAGAGCGG 2388
QY 1801 CAGAGGCAAGTACAGAGCTGACTGCGCGCGGAGCTGGCATGAAGAGAGCCCTTTTGA 1860
DB 2389 CAGAGGCAAGTACAGAGCTGACTGCGCGCGGAGCTGGCATGAAGAGAGCCCTTTTGA 2448
QY 1861 AAGCAGTTTAAAGCAGAGCTGCAAAATGAAATTTGAGAGAGCATCATGTCAAGAAC 1920
DB 2449 AAGCAGTTTAAAGCAGAGCTGCAAAATGAAATTTGAGAGAGCATCATGTCAAGAAC 2508
QY 1921 AGGTCAAGGAGAGAGCTGGGGGAAAGTGGGAGTCACTTAAGCTTTTTCGGGACAGATGAA 1980
DB 2509 AGGTCAAGGAGAGAGCTGGGGGAAAGTGGGAGTCACTTAAGCTTTTTCGGGACAGATGAA 2568
QY 1981 ATCATTGAGGTCTCC 1995

DB 2569 ATCATTGAGGTCTCC 2583
RESULT 23
AAH99685
ID AAH99685 strand; cDNA; 2966 BP.
AAH99685;
16-OCT-2001 (first entry)
Human protein encoding cDNA sequence SEQ ID NO:520.
Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
antihistaminic; antirheumatic; antiallergic; immunosuppressive;
antibacterial; endocrine; cardiac; central nervous system; virucide;
anti-HIV; fungicide; antimutagen; cardiovascular; antianemic; anaemia;
antiaggregant; haemostatic; vulnery; antitumor; osteopathic; eczema;
dermatological; antiallergic; antiasthmatic; antidiabetic; cytotoxic;
neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
genetic disease; haematopoietic disorder; platelet disorder; asthma;
thrombocytopenia; osteoporosis; severe combined immunodeficiency;
allergic rhinitis; diabetes; multiple sclerosis; depression;
Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
neurological disorder; ss.
Homo sapiens.
WO200153455-A2.
26-JUL-2001.
22-DEC-2000; 2000WO-US035017.
23-DEC-1999; 99US-00471275.
21-JAN-2000; 2000US-00488725.
25-APR-2000; 2000US-00552317.
(HYSE-) HYSEQ INC.
Tang YT, Liu C, Drmanac RT;
WPI; 2001-457603/49.
P-PDB; AAM25744.
Isolated human polynucleotides encoding polypeptides, useful for the
treatment and diagnosis of e.g. cancer, ulcers and HIV infection.
XX
XX
XX
PS Claim 1; Page 578; 1217pp; English.
CC AAH99166 to AAH99904 encode the human proteins given in AAM2525 to
CC AAM25963. The proteins can have activities based on the tissues and cells
CC they are expressed in, such as: antihistaminic; antirheumatic;
CC antiallergic; immunosuppressive; antibacterial; endocrine; cardiac;
CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
CC cardiovascular; antianemic; antiaggregant; haemostatic; vulnery;
CC antitumor; osteopathic; dermatological; antiallergic; antiasthmatic;
CC antidiabetic; cytotoxic; neuroprotective; antidepressant; nootropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC anaemia, platelet disorders, thrombocytopenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,

CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders
XX
SQ Sequence 2966 BP; 809 A; 735 C; 688 G; 733 T; 0 U; 1 Other;

Query Match 92.7%; Score 1849; DB 4; Length 2966;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1949; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
OY 45 GGTGGCTCTGCTGAAAGTGGAAAGGAAAAAGTCTGCTAATGTATGATGCGCGCATTTGT 104
Db 67 GGTGGCTCTGCTGAAAGTGGAAAGGAAAAAGTCTCTAATGTATGATGCGCGCATTTGT 126
OY 105 GGAATACAAATACATCCACATTTTGAAGCCATTATATCACTGCTCCAAAGTTTATGA 164
Db 127 GGAATACAAATACATCCACATTTTGAAGCCATTATATCACTGCTCCAAAGTTTATGA 186
OY 165 GCGAAGTTGGCAACAGACAAAGTTTAATTACAGAGCTCATCAGCATTCAGCGAAACA 224
Db 187 GCGAAGTTGGCAACAGACAAAGTTTAATTACAGAGCTCATCAGCATTCAGCGAAACA 246
OY 225 TAAAGTTGACATTGATGACAGTCAAGAGTTTGTATAGTATACAAAGCTCCCAAGATGT 284
Db 247 TAAAGTTGACATTGATGACAGTCAAGAGTTTGTATAGTATACAAAGCTCCCAAGATGT 306
OY 285 TGCCTCTCTCTCTCAGACTGTTTTCTCACTGTACTTCTGGTAACTGGAGAAAGACTT 344
Db 307 TGCCTCTCTCTCTCAGACTGTTTTCTCACTGTACTTCTGGTAACTGGAGAAAGACTT 366
OY 345 CAACCTCTGCTACCTGTTTGCAGAGTGGGTTTGCAGAGTTCTCTGTTGTTTCCGGCT 404
Db 367 CAACCTCTGCTACCTGTTTGCAGAGTGGGTTTGCAGAGTTCTCTGTTGTTTCCGGCT 426
OY 405 CTGTGAAGAAATTCACCTAGTCCCTACCTGACATTTCTCAAGCCCTTGGTTACCTGTTC 464
Db 427 CTGTGAAGAAATTCACCTAGTCCCTACCTGACATTTCTCAAGCCCTTGGTTACCTGTTC 486
OY 465 CAACATTGGGCCAACCCGGAATTTCTCCAACTTTTATCTTGGCTGCGACAGAGATGCT 524
Db 487 CAACATTGGGCCAACCCGGAATTTCTCCAACTTTTATCTTGGCTGCGACAGAGATGCT 546
OY 525 CAACAAGAGAGCTGATACAGACAAATGGGATTTGTGTAAATGCGACTATTAACCTG 584
Db 547 CAACAAGAGAGCTGATACAGACAAATGGGATTTGTGTAAATGCGACTATTAACCTG 606
OY 585 TCCAAAGCTGACTTATCCCGAGTCTCATTTCCGAGTCCGAGTCCGTAAGAAGACTT 644
Db 607 TCCAAAGCTGACTTATCCCGAGTCTCATTTCCGAGTCCGAGTCCGTAAGAAGACTT 666
OY 645 TTGTGAGAAATTTTGGCGGTGGTGGACAAATCAGTGAATTTCAATTGAGAAAGCAAAAG 704
Db 667 TTGTGAGAAATTTTGGCGGTGGTGGACAAATCAGTGAATTTCAATTGAGAAAGCAAAAG 726
OY 705 CTCGATGAGATGTTCTAGTGCACCTGTTTACTGAGATCTCCGCTCCGCCACCATGCG 764
Db 727 CTCGATGAGATGTTCTAGTGCACCTGTTTACTGAGATCTCCGCTCCGCCACCATGCG 786
OY 765 TATGCGCTACATCATGAGAGAGATGACATGCTTTAGTGAAGAGCTTACAGATTTGTGA 824
Db 787 TATGCGCTACATCATGAGAGAGATGACATGCTTTAGTGAAGAGCTTACAGATTTGTGA 846
OY 825 AGAAAAAGAGCTACTATATCTCCAACTTCAATTTTCTGGCCAACTCTGGAATATGA 884
Db 847 AGAAAAAGAGCTACTATATCTCCAACTTCAATTTTCTGGCCAACTCTGGAATATGA 906
OY 885 GAAAGAAGATTAGAACAGACGAGCTGAGCATCAGGGGCCAAAGCAAACTCAAGCTGCTGA 944
Db 907 GAAAGAAGATTAGAACAGACGAGCTGAGCATCAGGGGCCAAAGCAAACTCAAGCTGCTGA 966
OY 945 CCTGAGAAAGCCAAATGAACCTGTCTCTGCTGTCTCAAGAGGTGAGACAGAAAGCGAGAC 1004
Db 967 CCTGAGAAAGCCAAATGAACCTGTCTCTGCTGTCTCAAGAGGTGAGACAGAAAGCGAGAC 1026
```

```
OY 1005 GCCCTCACTCCACCCCTGTGCGGACTGTGTAACCTCAAGAGGACAGACAAAGGCCCT 1064
Db 1027 GCCCTCACTCCACCCCTGTGCGGACTGTGTAACCTCAAGAGGACAGACAAAGGCCCT 1086
OY 1065 GCATCCCGCAGAGGCGCCAGAGTGGCCAGAGTGGAGAGCCGTGCTGTATGAGAGACGCC 1124
Db 1087 GCATCCCGCAGAGGCGCCAGAGTGGCCAGAGTGGAGAGCCGTGCTGTATGAGAGACGCC 1146
OY 1125 GCTGTACAGGCGCTCAGTGGCTGACCTGTCCGACAGACAGGCTGGAAGACAGATTA 1184
Db 1147 GCTGTACAGGCGCTCAGTGGCTGACCTGTCCGACAGACAGGCTGGAAGACAGATTA 1206
OY 1185 GCTCAAGGCTTCTTCTCTCTGATATCAATCAGTTTCAATTCAGCCAGCATGCGACG 1244
Db 1207 GCTCAAGGCTTCTTCTCTCTGATATCAATCAGTTTCAATTCAGCCAGCATGCGACG 1266
OY 1245 ATCTTATACAGGCTTCTCTCATCAGAAAGATGCTTTGGAAATATACAAACCTTCACATAC 1304
Db 1267 ATCTTATACAGGCTTCTCTCATCAGAAAGATGCTTTGGAAATATACAAACCTTCACATAC 1326
OY 1305 TCTGATGAGAACCAACAAAGCTATGCAAGTTTCTCCCTGTTTCAGGAACATATGAGACAGAC 1364
Db 1327 TCTGATGAGAACCAACAAAGCTATGCAAGTTTCTCCCTGTTTCAGGAACATATGAGACAGAC 1386
OY 1365 TCCGAAACCAAGTCTCTGATAGAGAGAAAGCCAGATCCCAAGAACTGCAAGCCGACG 1424
Db 1387 TCCGAAACCAAGTCTCTGATAGAGAGAAAGCCAGATCCCAAGAAAGTGCAGAACCCGACG 1446
OY 1425 GCCTTCAGACAGCCAGACCAAGCATTTGCAATTCGGTCAGAAACCGACAGAGTGGCACCGC 1484
Db 1447 GCCTTCAGACAGCCAGACCAAGCATTTGCAATTCGGTCAGAAACCGACAGAGTGGCACCGC 1506
OY 1485 CCAAGATCCCTTTTATCTCACTGATCTGAATGGAGAGCGGTGAGGACAAATTAACAAC 1544
Db 1507 CCAAGATCCCTTTTATCTCACTGATCTGAATGGAGAGCGGTGAGGACAAATTAACAAC 1566
OY 1545 CAGCTTCCTTTTGGGCTTTTCCACAGCCAGACCACTTCAACGAAGTCTGTGGCTGGG 1604
Db 1567 CAGCTTCCTTTTGGGCTTTTCCACAGCCAGACCACTTCAACGAAGTCTGTGGCTGGG 1626
OY 1605 CTTTAAAGGCTGGGCACTCGGATATCTTTGGCCCCCAGACCTTACCCCTTCCCTGACAG 1664
Db 1627 CTTTAAAGGCTGGGCACTCGGATATCTTTGGCCCCCAGACCTTACCCCTTCCCTGACAG 1686
OY 1665 CAGTGTATATTTTCCACAGAGTCTGACACTTCTACTCTGCTCAAGCCATCTTACGAGG 1724
Db 1687 CAGTGTATATTTTCCACAGAGTCTGACACTTCTACTCTGCTCAAGCCATCTTACGAGG 1746
OY 1725 CAGTGCAGTTACTCTGCGCTTACAGCTGACGCGAGCTGCCACTTTGGGAGACCAAGTTTA 1784
Db 1747 CAGTGCAGTTACTCTGCGCTTACAGCTGACGCGAGCTGCCACTTTGGGAGACCAAGTTTA 1806
OY 1785 TTTGTGCGGAGGCGGACAGAGCCAAAGTGAACAGAGTGAACCTCGGCGGAGCTGACATGA 1844
Db 1807 TTTGTGCGGAGGCGGACAGAGCCAAAGTGAACAGAGTGAACCTCGGCGGAGCTGACATGA 1866
OY 1845 AGAGAGCCCTTTGAAAAGAGATTAAAGCAGAAAGCTGCAAAATGGAATTTTGGAGAGAG 1904
Db 1867 AGAGAGCCCTTTGAAAAGAGATTAAAGCAGAAAGCTGCAAAATGGAATTTTGGAGAGAG 1926
OY 1905 CATCATGTCAAGAAACAGGTACAGGGAAGAGCTGGGGAAGTGGGCAAGTCAGTCAAGCTT 1964
Db 1927 CATCATGTCAAGAAACAGGTACAGGGAAGAGCTGGGGAAGTGGGCAAGTCAGTCAAGCTT 1986
OY 1965 TTGCGGCAAGTGAATCATTTGAGGCTTCC 1995
Db 1987 TTGCGGCAAGTGAATCATTTGAGGCTTCC 2017
```

RESULT 24
AABD09492
ID AABD09492 standard; DNA; 2732 BP.
XX

AC AAD09492;
XX
DT 10-SEP-2001 (first entry)
XX
DE Human SGP002 phosphatase polypeptide encoding DNA.
XX
KM Human: SGP002 phosphatase polypeptide; phosphatase-related disease;
KM infection-related disorder; ocular disease; organ transplant rejection;
KM infection; diabetes; pain; sexual dysfunction; Alzheimer's disease;
KM metabolic disorder; haematopoietic cancer; mood disorder; cardiac;
KM Parkinson's disease; multiple sclerosis; amyotrophic lateral sclerosis;
KM cardiovascular disease; brain; neuronal-associated disease; dyskinesia;
KM attention disorder; cognition disorder; psychotic disorder; cytostatic;
KM neurological disorder; viricide; nootropic; cerebroprotective; therapy;
KM neuroprotective; antibacterial; antiviral; tranquilizer; antischistoma;
KM hypotensive; immunosuppressive; antipsoriatic; analgesic; hypertensive;
KM antifungal; dual specificity phosphatase; DSP; MAP kinase phosphatase;
KM MKP; mitraline; chromosome 12p11.1-p12.1; ds.
XX
OS Homo sapiens.
XX
FH Location/Qualifiers
FT CDS 538..2535
FT /tag= a
FT /product= "Human SGP002 phosphatase polypeptide"
XX
XX WO200146394-A2.
XX
PD 28-JUN-2001.
XX
PF 21-DEC-2000; 2000MO-US034736.
XX
PR 21-DEC-1999; 99US-0173255P.
XX PR 28-DEC-1999; 99US-0175766P.
XX PR 25-JAN-2000; 2000US-0178078P.
XX PR 31-JAN-2000; 2000US-0179301P.
XX
XX (SUGEN-) SUGEN INC.
XX
XX Plowman GD, Martinez R, Whyte D, Manning G, Sudarsanam S;
XX Hill RJ, Planagan P;
XX
XX MPI: 2001-418058/44.
XX P-PSDB; AAE04834.
XX
XX Novel phosphatase polypeptide useful for treating cancers, immune-related
XX diseases and disorders, cardiovascular disease, brain or neuronal-
XX associated diseases and metabolic disorders.
XX
XX Claim 29, Fig 1, 186pp, English.
XX
XX The present invention relates to phosphatase polypeptides, nucleotide
XX sequences encoding them, as well as various products and methods useful
XX for the diagnosis and treatment of various phosphatase-related diseases
XX and conditions. Substance that modulates the activity of phosphatase
XX polypeptide is used to treat immune-related diseases and disorders,
XX cardiovascular disease, brain or neuronal-associated diseases and
XX metabolic disorders, including cancers of tissues, cancers of
XX haematopoietic origin, diseases of central and peripheral nervous system,
XX Alzheimer's disease, Parkinson's disease, multiple sclerosis, amyotrophic
XX lateral sclerosis, viral infections, infections caused by prions,
XX bacteria and fungi, ocular diseases, diabetes, migraines, pain, sexual
XX dysfunction, mood disorders, attention disorders, cognitive disorders,
XX hypotension, hypertension, psychotic disorders, neurological disorders,
XX dyskinesias and organ transplant rejection. The present sequence is a DNA
XX encoding human SGP002 phosphatase polypeptide. This sequence is
XX classified as dual specificity phosphatase (DSP) and MAP kinase
XX phosphatase (MKP). SGP002 gene maps to chromosomal position 12p11.1-p12.1
XX
SQ Sequence 2732 BP; 710 A; 684 C; 686 G; 652 T; 0 U; 0 Other;

Query Match 92.3%; Score 1842; DB 4; Length 2732;
Best Local Similarity 99.8%; Pred. No. 0;

Matches 1992; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 ATGCCCATGAGATGTTGAACTCAAAATTGTACTGAGAGTTGGTGGCTCTGCTGGAA 60
DB 538 ATGCCCATGAGATGTTGAACTCAAAATTGTACTGAGAGTTGGTGGCTCTGCTGGAA 597
QY 61 AGTGAAGGAAAAAGTGTGCTAAATGATAGCCGGCCATTTGTGAAATACATATACATCC 120
DB 598 AGTGAAGGAAAAAGTGTGCTAAATGATAGCCGGCCATTTGTGAAATACATATACATCC 657
QY 121 CACATTTTGAAGCCATTAAATCAACCTGCTCAAGCTTATGAGCGAAGTTGCAACAG 180
DB 658 CACATTTTGAAGCCATTAAATCAACCTGCTCAAGCTTATGAGCGAAGTTGCAACAG 717
QY 181 GACAAAGTTTAATTACAGAGCTCATCCAGATTGCGCAAAATTAAGTTGACATTAT 240
DB 718 GACAAAGTTTAATTACAGAGCTCATCCAGATTGCGCAAAATTAAGTTGACATTAT 777
QY 241 TGCAGTCAGAGTTGATTTAAGATCAATCAAGTCCCAAGTGGCTCTGCTCTTCA 300
DB 778 TGCAGTCAGAGTTGATTTAAGATCAATCAAGTCCCAAGTGGCTCTGCTCTTCA 837
QY 301 GACTGTTTCTGACTGTACTTCTGGGTAATGAGAGAGCTTCAACTCTGTTCACTG 360
DB 838 GACTGTTTCTGACTGTACTTCTGGGTAATGAGAGAGCTTCAACTCTGTTCACTG 897
QY 361 CTTCAGAGTGGGTTTGTGAGTTCTCTGTTGTTTCCCTGGCTCTGTAAGAAATCC 420
DB 898 CTTCAGAGTGGGTTTGTGAGTTCTCTGTTGTTTCCCTGGCTCTGTAAGAAATCC 957
QY 421 ACTCTAGTCCCTACCTGATTTCTGAGCTTCTGCTTACCTGTTGCCAATTTGGCCAAAC 480
DB 958 ACTCTAGTCCCTACCTGATTTCTGAGCTTCTGCTTACCTGTTGCCAATTTGGCCAAAC 1017
QY 481 CCAATTCCTCCCAATCTTATCTGAGCTCCAGAGAGTGTCTTCAACAAGAGCTGAT 540
DB 1018 CCAATTCCTCCCAATCTTATCTGAGCTCCAGAGAGTGTCTTCAACAAGAGCTGAT 1077
QY 541 CAGCAGATGAGATGTTGTTATGTTAAATGCCAGTATACCTGTCCAAAGCTGACTTT 600
DB 1078 CAGCAGATGAGATGTTGTTATGTTAAATGCCAGTATACCTGTCCAAAGCTGACTTT 1137
QY 601 ATCCCGAGATCTCATTTCTGCGTGGCTGCTGGAATGAGAGCTTTGTGAGAAATTTTG 660
DB 1138 ATCCCGAGATCTCATTTCTGCGTGGCTGCTGGAATGAGAGCTTTGTGAGAAATTTTG 1197
QY 661 CCGTGGTGAAGAAATCAGTAGTTTCATTGAGAAAGCAAAAGCCCTCAATGATGTT 720
DB 1198 CCGTGGTGAAGAAATCAGTAGTTTCATTGAGAAAGCAAAAGCCCTCAATGATGTT 1257
QY 721 CTAGTCACATGTTTAACTGGGATCTCCGCTCCGCCACCATGCTATGCTTACATCATG 780
DB 1258 CTAGTCACATGTTTAACTGGGATCTCCGCTCCGCCACCATGCTATGCTTACATCATG 1317
QY 781 AAGAGATGAGATGCTTTAATGATGAGCTTACAGTTTGTGAGAAAGAAAGAAAGCTTACT 840
DB 1318 AAGAGATGAGATGCTTTAATGATGAGCTTACAGTTTGTGAGAAAGAAAGAAAGCTTACT 1377
QY 841 ATATCTCCAAACCTCAATTTTCTGGGCAATCTCTGAGCTATGAGAAAGATTAAGAAC 900
DB 1378 ATATCTCCAAACCTCAATTTTCTGGGCAATCTCTGAGCTATGAGAAAGATTAAGAAC 1437
QY 901 CAGACTGAGCATCAGGGCCAAAGCAAACTCAAGCTGCTCAGCTGAGAGCAAAAT 960
DB 1438 CAGACTGAGCATCAGGGCCAAAGCAAACTCAAGCTGCTCAGCTGAGAGCAAAAT 1497
QY 961 GAACCTGCTCCGCTGCTCTCAGAGGTGAGCAAGAAAGGAGACGCCCTTCACTCAACC 1020
DB 1498 GAACCTGCTCCGCTGCTCTCAGAGGTGAGCAAGAAAGGAGACGCCCTTCACTCAACC 1557
QY 1021 TGTGCCAATCTGCTACTCAGAGGCAAGAGCAAAAGGCGGTGATCCGCCAGCGTG 1080
DB 1558 TGTGCCAATCTGCTACTCAGAGGCAAGAGCAAAAGGCGGTGATCCGCCAGCGTG 1617

```

QY 1081 CCAGAGGTGCCAGGCGTCCGCTGTTAGAGACAGCCCGCTGTTACAGCGCTC 1140
DB 1618 CCCAGAGTCCCAAGCGTGGAGCCGTGGTTAGAGAGAGCCCGCTGTTACAGCGCTC 1677
QY 1141 AGTGGCTGACCTGTGCGCAGACAGGTGAGAGACACAAATTAAGCTCAAGGTTCCCTC 1200
DB 1678 AGTGGCTGACCTGTGCGCAGACAGGTGAGAGACACAAATTAAGCTCAAGGTTCCCTC 1737
QY 1201 TCTCGGATATCAATTCAGTTTCAATTCAGCAGCATGCGACAGCATTCCTTACATGCTTC 1260
DB 1738 TCTCGGATATCAATTCAGTTTCAATTCAGCAGCATGCGACAGCATTCCTTACATGCTTC 1797
QY 1261 TCTCGATCAGAAAGATGCTTTGGAAATACAAACCTTCACATCTGGAAAGGGAACCAAC 1320
DB 1798 TCTCGATCAGAAAGATGCTTTGGAAATACAAACCTTCACATCTGGAAAGGGAACCAAC 1857
QY 1321 AAGCTATGCCAGTTCTCCCTGTTGAGAACTATCGGAGCAGACTCCGAAACAGTCTC 1380
DB 1858 AAGCTATGCCAGTTCTCCCTGTTGAGAACTATCGGAGCAGACTCCGAAACAGTCTC 1917
QY 1381 GATAAGAGAGAGCAGACATCCCAAGAGCTGAGACCCGCAAGCTTCAGACAGCAG 1440
DB 1918 GATAAGAGAGAGCAGACATCCCAAGAGCTGAGACCCGCAAGCTTCAGACAGCAG 1977
QY 1441 AGCAAGCGATTGCAATTCGCTGAGAACAGCAGCAGTGGCAGCCGCGAGAGTCCCTTTA 1500
DB 1978 AGCAAGCGATTGCAATTCGCTGAGAACAGCAGCAGTGGCAGCCGCGAGAGTCCCTTTA 2037
QY 1501 TCTGACATGCAATCGAAATGGAGAGCTGAGAGACATTAACAACACAGCTTCCTTTTCGCG 1560
DB 2038 TCTGACATGCAATCGAAATGGAGAGCTGAGAGACATTAACAACACAGCTTCCTTTTCGCG 2097
QY 1561 CTTCACACAGCCAGCAGACCTTCACAGAGTGTGCTGGGCTTAAAGGCTGGCAC 1620
DB 2098 CTTCACACAGCCAGCAGACCTTCACAGAGTGTGCTGGGCTTAAAGGCTGGCAC 2157
QY 1621 TCGGATATCTTTGGGCCCCCAGACCTCTAACCCTTCCCTGACAGCAGCTGTATTTTGGC 1680
DB 2158 TCGGATATCTTTGGGCCCCCAGACCTCTAACCCTTCCCTGACAGCAGCTGTATTTTGGC 2217
QY 1681 ACAGAGCTCTCAGACCTCTCTGCTGCTGACCATCTACGAGAGGCTGCCAGTACTCT 1740
DB 2218 ACAGAGCTCTCAGACCTCTCTGCTGCTGACCATCTACGAGAGGCTGCCAGTACTCT 2277
QY 1741 GCTTACAGCTGCAAGCCAGCTGCGCACTTGGAGAGACCAAGTCTATTTCTGGCAGAGCG 1800
DB 2278 GCTTACAGCTGCAAGCCAGCTGCGCACTTGGAGAGACCAAGTCTATTTCTGGCAGAGCG 2337
QY 1801 CAGAAGCCAGTGAACAGAGCTGATCTGCGCGCGAGCTGGCATGAAAGAGAGCCCTTTGAA 1860
DB 2338 CAGAAGCCAGTGAACAGAGCTGATCTGCGCGCGAGCTGGCATGAAAGAGAGCCCTTTGAA 2397
QY 1861 AAGCAGTTTAAACGAGAGCTGCCAAATGGAATTTGAGAGAGATCAATGTCAGAGAAC 1920
DB 2398 AAGCAGTTTAAACGAGAGCTGCCAAATGGAATTTGAGAGAGATCAATGTCAGAGAAC 2457
QY 1921 AGGTCAAGGAGAGAGCTGGGAAAAGTGGGCACTGCTTATTTCTGGGCGAGATGAA 1980
DB 2458 AGGTCAAGGAGAGAGCTGGGAAAAGTGGGCACTGCTTATTTCTGGGCGAGATGAA 2517
QY 1981 ATCATTTGAGGTCTCC 1995
DB 2518 ATCATTTGAGGTCTCC 2532

```

RESULT 25
ABNS9704
ID ABNS9704 standard; cDNA; 3104 BP.

AC ABNS9704;
XX
DT 28-JUN-2002 (first entry)

```

XX DE Novel human coding sequence SEQ ID NO: 115.
XX KW Human; anti-neurotic; vulnerability; anti-inflammatory; immunomodulator;
XX KW anti-infectivity; cerebroprotective; cytoskeletal; rheumatic; gene therapy;
XX KW neuroprotective; anti-parkinsonian; protein therapy; EST;
XX KW expressed sequence tag; gene; ss.
XX OS Homo sapiens.
XX PN M020022660-A2.
XX PD 21-MAR-2002.
XX PF 10-SEP-2001; 2001WO-US026015.
XX PR 11-SEP-2000; 2000US-00659671.
XX PA (HSP-) HYSEQ INC.
XX PI Tang YF, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
XX PI Xue AJ, Yang Y, Wehrman T, Drmanac RT,
XX DR WPI; 2002-292408/33.
XX P-PSDB; ABB97291.
XX PT An isolated polynucleotide for treating diseases associated with its
XX encoded polypeptide such as cancer and multiple sclerosis.
XX PS Claim 1, SEQ ID NO 115; 509bp; English.
XX CC The present invention provides the protein and coding sequences of 444
XX CC novel human proteins. These were isolated from expressed sequences tags
XX CC (ESTs). They can be used to stimulate cell growth, to regulate
XX CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
XX CC e.g. in burn treatment, to regulate the immune system e.g. to treat
XX CC multiple sclerosis, to regulate the immune system e.g. to treat
XX CC infertility, to regulate haemostasis or thrombolysis e.g. to treat
XX CC and cancer, to screen for drugs, to treat inflammatory conditions e.g.
XX CC rheumatoid arthritis, and to treat nervous system disorders e.g.
XX CC Parkinson's disease. The present sequence is a coding sequence of the
XX CC invention
XX SQ Sequence 3104 BP; 845 A; 772 C; 718 G; 769 T; 0 U; 0 Other;
QY 1 ATGGCCCATGAGATGATTTGGAATCAAAATTTGTAATCTGAGAGGTTGGTCTGCTGGAA 60
DB 159 ATGGCCCATGAGATGATTTGGAATCAAAATTTGTAATCTGAGAGGTTGGTCTGCTGGAA 218
QY 61 AGTGAACCGAAAAAGTGTCTGTAATTTGATAGCGCGCCATTGTGTGGAATACATCATCC 120
DB 219 AGTGAACCGAAAAAGTGTCTGTAATTTGATAGCGCGCCATTGTGTGGAATACATCATCC 278
QY 121 CACATTTTGAAGCCATTAATTAATCACTGCTCAAGCTTATGAAGGAAAGTTGCAACG 180
DB 279 CACATTTTGAAGCCATTAATTAATCACTGCTCAAGCTTATGAAGGAAAGTTGCAACG 338
QY 181 GACAAAGTGTAAATTCAGAGCTCATCCAGACTTACAGGGAACATTAAGTTGACATTCAT 240
DB 339 GACAAAGTGTAAATTCAGAGCTCATCCAGACTTACAGGGAACATTAAGTTGACATTCAT 398
QY 241 TGCAGTCAGAGGTTGATTTAGATCAAAAGCTCCCAAGATGTTGCTCTCTCTTCA 300
DB 399 TGCAGTCAGAGGTTGATTTAGATCAAAAGCTCCCAAGATGTTGCTCTCTCTTCA 458
QY 301 GACTGTTTCTCAGTGTACTTCTGGTAAACTGAGGAAAGAGTTTAACTTCTGTTCACTG 360
DB 459 GACTGTTTCTCAGTGTACTTCTGGTAAACTGAGGAAAGAGTTTAACTTCTGTTCACTG 518

```

QY 361 CTTCAGAGTGGGTTTGGCTGAGTTCTCTCGTTGTTCCCTGGCCCTGTGTAAGGAAAATCC 420
 DB 519 CTTCAGAGTGGGTTTGGCTGAGTTCTCTCGTTGTTCCCTGGCCCTGTGTAAGGAAAATCC 578
 QY 421 ACTCTAGTCCCTACCTGCAATTTCTCAGCCCTGCTTACCTGTGGCAACATTTGGCCAAACC 480
 DB 579 ACTCTAGTCCCTACCTGCAATTTCTCAGCCCTGCTTACCTGTGGCAACATTTGGCCAAACC 638
 QY 481 CGAATTTCTCCCAATCTTTATCTTGGCTGCCAGCAGATGTCTCTCAACAGAGCTGATA 540
 DB 639 CGAATTTCTCCCAATCTTTATCTTGGCTGCCAGCAGATGTCTCTCAACAGAGCTGATA 698
 QY 541 CAGCAGAAATGGATGGTATGTGTAAATGCCAGTATACCTGTCCAAAGCCGATCTT 600
 DB 699 CAGCAGAAATGGATGGTATGTGTAAATGCCAGTATACCTGTCCAAAGCCGATCTT 758
 QY 601 ATCCCGAGTCTCATTTCTGCGTGTGCTGTGATGACAGCTTTGTGAAAATTTTG 660
 DB 759 ATCCCGAGTCTCATTTCTGCGTGTGCTGTGATGACAGCTTTGTGAAAATTTTG 818
 QY 661 CCGTGTGTGCAAAATAGTAGATTTTATGAGAAAACAAAGCCCTGCATGATGTT 720
 DB 819 CCGTGTGTGCAAAATAGTAGATTTTATGAGAAAACAAAGCCCTGCATGATGTT 878
 QY 721 CTAGTGCACTGTTAGCTGGGATCTCCGCTCCGCCACATGCTATGCTCATCATG 780
 DB 879 CTAGTGCACTGTTAGCTGGGATCTCCGCTCCGCCACATGCTATGCTCATCATG 938
 QY 781 AAGAGATGACATGTCTTATGATGAACTTACAGATTTGTGAAAAGAAAAGACCTACT 840
 DB 939 AAGAGATGACATGTCTTATGATGAACTTACAGATTTGTGAAAAGAAAAGACCTACT 998
 QY 841 ATATCTCCAACTTCAATTTTCTGGGCAACTCTGAGCTATGAGAAAAGATTAAGAAC 900
 DB 999 ATATCTCCCAACTTCAATTTTCTGGGCAACTCTGAGCTATGAGAAAAGATTAAGAAC 1058
 QY 901 CAACTGAGACATCAGGGCCAAAGAGCAACTCAAGCTGTGACCTGGAGAACCAAT 960
 DB 1059 CAACTGAGACATCAGGGCCAAAGAGCAACTCAAGCTGTGACCTGGAGAACCAAT 1118
 QY 961 GAACTGTCCCTCTGTCTCAGAGGTGACAGAAAAGCAGACGCCCTCACTGCACCC 1020
 DB 1119 GAACTGTCCCTCTGTCTCAGAGGTGACAGAAAAGCAGACGCCCTCACTGCACCC 1178
 QY 1021 TGTGCGCACTCTGTCTCAGAGGTGACAGAAAAGCAGACGCCCTCACTGCACCC 1080
 DB 1179 TGTGCGCACTCTGTCTCAGAGGTGACAGAAAAGCAGACGCCCTCACTGCACCC 1238
 QY 1081 CCCAGGTGCGCAGCGTGACGCGTGTGTAGAGGACGCCGCTGTGACAGGGGCTC 1140
 DB 1239 CCCAGGTGCGCAGCGTGACGCGTGTGTAGAGGACGCCGCTGTGACAGGGGCTC 1298
 QY 1141 AGTGGGTGCACTGTCTCCGACAGAGCTGGAAGACAGCAATTAAGCTCAAGCTTCTTC 1200
 DB 1299 AGTGGGTGCACTGTCTCCGACAGAGCTGGAAGACAGCAATTAAGCTCAAGCTTCTTC 1358
 QY 1201 TCTCTGATATCAATATGATTTATATATCAAGCAGATGAGCAGATCTTACATGCTTC 1260
 DB 1359 TCTCTGATATCAATATGATTTATATATCAAGCAGATGAGCAGATCTTACATGCTTC 1418
 QY 1261 TCCTCATCGAAGATGTTTGAATCTAATAAATTCTTCCATCTCTGATGAGGACCAAC 1320
 DB 1419 TCCTCATCGAAGATGTTTGAATCTAATAAATTCTTCCATCTCTGATGAGGACCAAC 1478
 QY 1321 AAGCTATGCAAGTTTCTCCCTGTTCAGAACTATCGAGCAGACTCCGAAACCACTGCT 1380
 DB 1479 AAGCTATGCAAGTTTCTCCCTGTTCAGAACTATCGAGCAGACTCCGAAACCACTGCT 1538
 QY 1381 GATTAAGAGAGACCGACATCTCCCAAGAGCTGACACCGCCAGGCTTCAAGACGCCAG 1440
 DB 1539 GATTAAGAGAGACCGACATCTCCCAAGAGCTGACACCGCCAGGCTTCAAGACGCCAG 1598
 QY 1441 AGCAAGCATTTGATTCGATCAAAACAGACAGAGTGGCCAGCCGCAAGTCTCTTTA 1500

DB 1599 AGCAAGCATTTGATTTGGATCGAGAACACAGAGAGTGGACACGCCCCAGAGTCTCTTTA 1658
 QY 1501 TCTCCACTGATCGAAGTGGAGCGGTGAGAGACATTTACACACAGCTCTCTTTGCGC 1560
 DB 1659 TCTCCACTGATCGAAGTGGAGCGGTGAGAGACATTTACACACAGCTCTCTTTGCGC 1718
 QY 1561 CTTTCCACAGACAGACACCTCAGAAAGTCTGCTGGCTGGGCTTTAAGGCTGGCAC 1620
 DB 1719 CTTTCCACAGACAGACACCTCAGAAAGTCTGCTGGGCTGGGCTTTAAGGCTGGCAC 1778
 QY 1621 TCAGATATCTTGGCCCCCAGACCTTACCCCTTCCCTGACACAGACCTGTATTTTGGC 1680
 DB 1779 TCAGATATCTTGGCCCCCAGACCTTACCCCTTCCCTGACACAGACCTGTATTTTGGC 1838
 QY 1681 ACAGAGTCTCACACTTCTACTCTGCTCAGACCATTTACGAGAGCAGTGCATTACTT 1740
 DB 1839 ACAGAGTCTCACACTTCTACTCTGCTCAGACCATTTACGAGAGCAGTGCATTACTT 1898
 QY 1741 GCCTACAGCTGACGCCAGCTGCCACTTGTGGAGACCAAGTCTATTCTGTGGCAGGCG 1800
 DB 1899 GCCTACAGCTGACGCCAGCTGCCACTTGTGGAGACCAAGTCTATTCTGTGGCAGGCG 1958
 QY 1801 CAGAACCAAGTGAACAGACTGACTGCGCGCGAGCTGACATGAAGAGAGCCCTTTGAA 1860
 DB 1959 CAGAACCAAGTGAACAGACTGACTGCGCGCGAGCTGACATGAAGAGAGCCCTTTGAA 2018
 QY 1861 AAGCAGTTTAAACGAGAACTGCCAATGAAATTTGAGAGAGCATCTGTGAGAAC 1920
 DB 2019 AAGCAGTTTAAACGAGAACTGCCAATGAAATTTGAGAGAGCATCTGTGAGAAC 2078
 QY 1921 AGGTACGGGAAAGAGTGGGAAAGTGGGCACTCACTTCTTGGGACAGATGAA 1980
 DB 2079 AGGTACGGGAAAGAGTGGGAAAGTGGGCACTCACTTCTTGGGACAGATGAA 2138
 QY 1981 ATCATTGAGTCTCC 1995
 DB 2139 ATCATTGAGTCTCC 2153

RESULT 26
 ADAS3105
 ID ADAS3105 standard; cDNA; 2102 BP.
 XX
 AC ADAS3105;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE Human coding sequence, SEQ ID 673.
 XX
 KW Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nootropic;
 KW Gene Therapy; human; secretory protein; membrane proteins; cancer;
 KW Inflammatory disease; osteoporosis; neurological disease; gene; ss.
 OS Homo sapiens.
 PN EP1293569-A2.
 PD 19-MAR-2003.
 PF 21-MAR-2002; 2002EP-00006586.
 PR 14-SEP-2001; 2001JP-00328381.
 PR 24-JAN-2002; 2002US-0350435P.
 XX
 PA (HELI-) HELIX RES INST.
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.
 PI Tsogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,
 PI Yamamoto U, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,
 PI Seki N, Yoshikawa T, Otsuka M, Nagahara K, Masuho Y,
 XX
 DR WPI; 2003-395539/38.

DR F-PSDB; ADA54744.
XX New polynucleotides encoding full-length polypeptides, e.g. secretory
PT and/or membrane proteins, useful for developing medicines for diseases in
PT which the gene is involved, or as target molecules for gene therapy.
XX
PS Claim 1; SEQ ID NO 673; 205bp; English.
XX
CC The present invention relates to novel human secretory or membrane
CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
CC ADA54071). The coding sequences are useful in the gene therapy of
CC diseases caused by abnormalities of the proteins, e.g. cancer,
CC inflammatory diseases, osteoporosis or neurological disease.
XX
SQ Sequence 2102 BP; 542 A; 546 C; 510 G; 504 T; 0 U; 0 Other;

Query Match 89.8%; Score 1791; DB 7; Length 2102;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1991; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 ATGGCCCATGAGATGATGAACTCAATTGTTACTGAGAGGTGGTGGCTCTGCTGGAA 60
Db ATGGCCCATGAGATGATGAACTCAATTGTTACTGAGAGGTGGTGGCTCTGCTGGAA 115
QY 61 AGTGAACGGAAGAGTCTGCTAATTGATAGCCGCAATTGTGGAATACAATACATCC 120
Db AGTGAACGGAAGAGTCTGCTAATTGATAGCCGCAATTGTGGAATACAATACATCC 175
QY 121 CACATTTTGGAAAGCATTAAATCACTGCTCCAGCTTATGAAAGCAAGTTGCAACAG 180
Db CACATTTTGGAAAGCATTAAATCACTGCTCCAGCTTATGAAAGCAAGTTGCAACAG 235
QY 176 CACATTTTGGAAAGCATTAAATCACTGCTCCAGCTTATGAAAGCAAGTTGCAACAG 235
Db CACATTTTGGAAAGCATTAAATCACTGCTCCAGCTTATGAAAGCAAGTTGCAACAG 295
QY 236 GACAAAGTGTAAATTAACAAGCTCATCAGCATTCAGCGAAACATAAGTTGACATTGAT 295
Db GACAAAGTGTAAATTAACAAGCTCATCAGCATTCAGCGAAACATAAGTTGACATTGAT 295
QY 241 TGCAGTGAAGAGGTGATGATTAACATCAAGCTCCCAAGATGTTGCTCTCTCTTCA 300
Db TGCAGTGAAGAGGTGATGATTAACATCAAGCTCCCAAGATGTTGCTCTCTCTTCA 355
QY 296 TGCAGTGAAGAGGTGATGATTAACATCAAGCTCCCAAGATGTTGCTCTCTCTTCA 355
Db TGCAGTGAAGAGGTGATGATTAACATCAAGCTCCCAAGATGTTGCTCTCTCTTCA 360
QY 301 GACTGTTTCTCACTGACTTCTGGGTAACTGGAAGAGCTTCACTCTGTTCACTG 360
Db GACTGTTTCTCACTGACTTCTGGGTAACTGGAAGAGCTTCACTCTGTTCACTG 415
QY 356 GACTGTTTCTCACTGACTTCTGGGTAACTGGAAGAGCTTCACTCTGTTCACTG 415
Db GACTGTTTCTCACTGACTTCTGGGTAACTGGAAGAGCTTCACTCTGTTCACTG 475
QY 361 CTTCGAGTGGGTTTGTGAGTCTCTGTTTTCCTGGCTCTGTAAGAAATTC 420
Db CTTCGAGTGGGTTTGTGAGTCTCTGTTTTCCTGGCTCTGTAAGAAATTC 475
QY 416 CTTCGAGTGGGTTTGTGAGTCTCTGTTTTCCTGGCTCTGTAAGAAATTC 475
Db CTTCGAGTGGGTTTGTGAGTCTCTGTTTTCCTGGCTCTGTAAGAAATTC 480
QY 421 ACTCTAGTCCCTAAGTCTGCAATTTCTGAGCTTGTACTGTTCCCAACTTGGCCAA 480
Db ACTCTAGTCCCTAAGTCTGCAATTTCTGAGCTTGTACTGTTCCCAACTTGGCCAA 535
QY 476 ACTCTAGTCCCTAAGTCTGCAATTTCTGAGCTTGTACTGTTCCCAACTTGGCCAA 535
Db ACTCTAGTCCCTAAGTCTGCAATTTCTGAGCTTGTACTGTTCCCAACTTGGCCAA 540
QY 481 CGAATTTCTCCCAATCTTATCTTGAGCTCCAGCGAGATGCTCTCAAGAGAGCTGATA 540
Db CGAATTTCTCCCAATCTTATCTTGAGCTCCAGCGAGATGCTCTCAAGAGAGCTGATA 595
QY 536 CTAAATCTTCCCAATCTTATCTTGAGCTCCAGCGAGATGCTCTCAAGAGAGCTGATA 595
Db CTAAATCTTCCCAATCTTATCTTGAGCTCCAGCGAGATGCTCTCAAGAGAGCTGATA 600
QY 541 CAGCAAAATGGAATGTTGTTAATGTCAGCTATACCTGTGCCAAGCTGACTTT 600
Db CAGCAAAATGGAATGTTGTTAATGTCAGCTATACCTGTGCCAAGCTGACTTT 655
QY 596 CAGCAAAATGGAATGTTGTTAATGTCAGCTATACCTGTGCCAAGCTGACTTT 655
Db CAGCAAAATGGAATGTTGTTAATGTCAGCTATACCTGTGCCAAGCTGACTTT 660
QY 601 ATCCCGAGTCTCATTTTCTGCGTGGCTGAGTGAAGAGCTTTTGGAAATTTTGG 660
Db ATCCCGAGTCTCATTTTCTGCGTGGCTGAGTGAAGAGCTTTTGGAAATTTTGG 715
QY 656 ATCCCGAGTCTCATTTTCTGCGTGGCTGAGTGAAGAGCTTTTGGAAATTTTGG 715
Db ATCCCGAGTCTCATTTTCTGCGTGGCTGAGTGAAGAGCTTTTGGAAATTTTGG 720
QY 661 CCGTGTGTTGACAAATCAGTAGATTTCATTGAGAAAGCAAAAGCTCCCAATGATGTT 720
Db CCGTGTGTTGACAAATCAGTAGATTTCATTGAGAAAGCAAAAGCTCCCAATGATGTT 775
QY 716 CCGTGTGTTGACAAATCAGTAGATTTCATTGAGAAAGCAAAAGCTCCCAATGATGTT 775
Db CCGTGTGTTGACAAATCAGTAGATTTCATTGAGAAAGCAAAAGCTCCCAATGATGTT 780
QY 721 CTAGTGAATGTTGTTGCTGGGATCTCCCGCTCCGCAACATGCTATGCTCATCATATG 780
Db CTAGTGAATGTTGTTGCTGGGATCTCCCGCTCCGCAACATGCTATGCTCATCATATG 835
QY 776 CTAGTGAATGTTGTTGCTGGGATCTCCCGCTCCGCAACATGCTATGCTCATCATATG 835
Db CTAGTGAATGTTGTTGCTGGGATCTCCCGCTCCGCAACATGCTATGCTCATCATATG 840
QY 781 AAGAGATGACATGCTTTAGATGAAGCTTACAGATTGTGAAAGAAAGCACTACT 840

Db AAGAGATGACATGCTTTAGATGAAGCTTACAGATTGTGAAAGAAAGCACTACT 895
QY 836 AAGAGATGACATGCTTTAGATGAAGCTTACAGATTGTGAAAGAAAGCACTACT 895
QY 841 ATATCTCCCAATCTTCAATTTTCTGGGCCCACTCTGACATATGAGAAGATTAAAGAC 900
Db ATATCTCCCAATCTTCAATTTTCTGGGCCCACTCTGACATATGAGAAGATTAAAGAC 955
QY 896 ATATCTCCCAATCTTCAATTTTCTGGGCCCACTCTGACATATGAGAAGATTAAAGAC 955
Db ATATCTCCCAATCTTCAATTTTCTGGGCCCACTCTGACATATGAGAAGATTAAAGAC 960
QY 901 CAGACTGAGCATCAGGGCCAAAGAGCAAACTCAAGCTGTGCACCTGAGAAAGCAAT 960
Db CAGACTGAGCATCAGGGCCAAAGAGCAAACTCAAGCTGTGCACCTGAGAAAGCAAT 1015
QY 956 CAGACTGAGCATCAGGGCCAAAGAGCAAACTCAAGCTGTGCACCTGAGAAAGCAAT 1015
Db CAGACTGAGCATCAGGGCCAAAGAGCAAACTCAAGCTGTGCACCTGAGAAAGCAAT 1020
QY 961 GAACTGTCTCTGCTGTCTCAGAGGTGAGCAAAAGCAAGAGCCCTCTGACATCC 1020
Db GAACTGTCTCTGCTGTCTCAGAGGTGAGCAAAAGCAAGAGCCCTCTGACATCC 1075
QY 1021 TGTGCGAATCTGCTACCTCAGAGGACAGAGCAAAAGCCCTGTGATCCCGCAGGCTG 1080
Db TGTGCGAATCTGCTACCTCAGAGGACAGAGCAAAAGCCCTGTGATCCCGCAGGCTG 1135
QY 1076 TGTGCGAATCTGCTACCTCAGAGGACAGAGCAAAAGCCCTGTGATCCCGCAGGCTG 1135
Db TGTGCGAATCTGCTACCTCAGAGGACAGAGCAAAAGCCCTGTGATCCCGCAGGCTG 1140
QY 1081 CCCAGCGTCCAGCGTCCAGCGCTGCTGTTAAGAGACAGCCCGCTGTACAGGCTC 1140
Db CCCAGCGTCCAGCGTCCAGCGCTGCTGTTAAGAGACAGCCCGCTGTACAGGCTC 1195
QY 1136 CCCAGCGTCCAGCGTCCAGCGCTGCTGTTAAGAGACAGCCCGCTGTACAGGCTC 1195
Db CCCAGCGTCCAGCGTCCAGCGCTGCTGTTAAGAGACAGCCCGCTGTACAGGCTC 1200
QY 1141 AGTGGGCTGCACTGTCTGCGAGACAGAGCTGGAAGACAGCAATTAAGTTCTTC 1200
Db AGTGGGCTGCACTGTCTGCGAGACAGAGCTGGAAGACAGCAATTAAGTTCTTC 1255
QY 1196 AGTGGGCTGCACTGTCTGCGAGACAGAGCTGGAAGACAGCAATTAAGTTCTTC 1255
Db AGTGGGCTGCACTGTCTGCGAGACAGAGCTGGAAGACAGCAATTAAGTTCTTC 1260
QY 1201 TCTCTGATTAATCAATCAGTTTCAATTTCAAGCCAGATGAGAGATCCCTTAAGGCTTC 1260
Db TCTCTGATTAATCAATCAGTTTCAATTTCAAGCCAGATGAGAGATCCCTTAAGGCTTC 1315
QY 1256 TCTCTGATTAATCAATCAGTTTCAATTTCAAGCCAGATGAGAGATCCCTTAAGGCTTC 1315
Db TCTCTGATTAATCAATCAGTTTCAATTTCAAGCCAGATGAGAGATCCCTTAAGGCTTC 1320
QY 1261 TCTCTATCAGAAAGTCTTTGGAATTAATCAAACTTCACTCTGATGAGAGCAAC 1320
Db TCTCTATCAGAAAGTCTTTGGAATTAATCAAACTTCACTCTGATGAGAGCAAC 1375
QY 1316 TCTCTATCAGAAAGTCTTTGGAATTAATCAAACTTCACTCTGATGAGAGCAAC 1375
Db TCTCTATCAGAAAGTCTTTGGAATTAATCAAACTTCACTCTGATGAGAGCAAC 1380
QY 1321 AAGCTATGCGAGTCTCCCGTTCAGAGAACTATCGAGAGACCTCCGAAACCACTGCT 1380
Db AAGCTATGCGAGTCTCCCGTTCAGAGAACTATCGAGAGACCTCCGAAACCACTGCT 1435
QY 1376 AAGCTATGCGAGTCTCCCGTTCAGAGAACTATCGAGAGACCTCCGAAACCACTGCT 1435
Db AAGCTATGCGAGTCTCCCGTTCAGAGAACTATCGAGAGACCTCCGAAACCACTGCT 1440
QY 1381 GATTAAGAGAGAGCAGACATCCCAAGAGCTCAGACCGCCTTCAAGACAG 1440
Db GATTAAGAGAGAGCAGACATCCCAAGAGCTCAGACCGCCTTCAAGACAG 1495
QY 1436 GATTAAGAGAGAGCAGACATCCCAAGAGCTCAGACCGCCTTCAAGACAG 1495
Db GATTAAGAGAGAGCAGACATCCCAAGAGCTCAGACCGCCTTCAAGACAG 1500
QY 1441 AGCAAGGATTTGATTTGCTGCTCAAGAACAGAGAGTGCACCGCCAGAGTCCCTTTTA 1500
Db AGCAAGGATTTGATTTGCTGCTCAAGAACAGAGAGTGCACCGCCAGAGTCCCTTTTA 1555
QY 1496 AGCAAGGATTTGATTTGCTGCTCAAGAACAGAGAGTGCACCGCCAGAGTCCCTTTTA 1555
Db AGCAAGGATTTGATTTGCTGCTCAAGAACAGAGAGTGCACCGCCAGAGTCCCTTTTA 1560
QY 1501 TCTCACTGATGAGAGTGGAGAGCTGAGAGCAATTTACACACAGAGCTTCCCTTTTGG 1560
Db TCTCACTGATGAGAGTGGAGAGCTGAGAGCAATTTACACACAGAGCTTCCCTTTTGG 1615
QY 1556 TCTCACTGATGAGAGTGGAGAGCTGAGAGCAATTTACACACAGAGCTTCCCTTTTGG 1615
Db TCTCACTGATGAGAGTGGAGAGCTGAGAGCAATTTACACACAGAGCTTCCCTTTTGG 1620
QY 1561 CTTTCCACACAGCAGACAGACCTTCAAGAACTGCTGAGCTGAGGCTTAAAGGCTGGAC 1620
Db CTTTCCACACAGCAGCAGACCTTCAAGAACTGCTGAGCTGAGGCTTAAAGGCTGGAC 1675
QY 1616 CTTTCCACACAGCAGCAGACCTTCAAGAACTGCTGAGCTGAGGCTTAAAGGCTGGAC 1675
Db CTTTCCACACAGCAGCAGACCTTCAAGAACTGCTGAGCTGAGGCTTAAAGGCTGGAC 1680
QY 1621 TCGGATATCTTGGAGCTCCCAAGCTTACCTTCCCTGACAGAGCTGATTTTTC 1680
Db TCGGATATCTTGGAGCTCCCAAGCTTACCTTCCCTGACAGAGCTGATTTTTC 1735
QY 1676 TCGGATATCTTGGAGCTCCCAAGCTTACCTTCCCTGACAGAGCTGATTTTTC 1735
Db TCGGATATCTTGGAGCTCCCAAGCTTACCTTCCCTGACAGAGCTGATTTTTC 1740
QY 1681 ACAGAGTCTCAACTTCTACTCTGCTCAGACATCTACAGAGGAGGAGGCTTACTCT 1740
Db ACAGAGTCTCAACTTCTACTCTGCTCAGACATCTACAGAGGAGGAGGCTTACTCT 1795
QY 1736 ACAGAGTCTCAACTTCTACTCTGCTCAGACATCTACAGAGGAGGAGGCTTACTCT 1795
Db ACAGAGTCTCAACTTCTACTCTGCTCAGACATCTACAGAGGAGGAGGCTTACTCT 1800
QY 1741 GCTTACAGCTGAGAGCAGCTGCGCCTTGGGAGACCAAGTCTTCTGTCGAGGCGG 1800
Db GCTTACAGCTGAGAGCAGCTGCGCCTTGGGAGACCAAGTCTTCTGTCGAGGCGG 1855
QY 1796 GCTTACAGCTGAGAGCAGCTGCGCCTTGGGAGACCAAGTCTTCTGTCGAGGCGG 1855
Db GCTTACAGCTGAGAGCAGCTGCGCCTTGGGAGACCAAGTCTTCTGTCGAGGCGG 1860
QY 1801 CAGAAAGCAAGTGAAGAGAGCTGACTGCGGAGGAGCTGGCATGAAAGAGCCCTTTGAA 1860
Db CAGAAAGCAAGTGAAGAGAGCTGACTGCGGAGGAGCTGGCATGAAAGAGCCCTTTGAA 1915
QY 1856 CAGAAAGCAAGTGAAGAGAGCTGACTGCGGAGGAGCTGGCATGAAAGAGCCCTTTGAA 1915
Db CAGAAAGCAAGTGAAGAGAGCTGACTGCGGAGGAGCTGGCATGAAAGAGCCCTTTGAA 1920
QY 1861 AAGCATTTAAACGAGAGAGCTGCCAAATGGAATTTGAGAGAGCATCATGTCAGAAAC 1920
Db AAGCATTTAAACGAGAGAGCTGCCAAATGGAATTTGAGAGAGCATCATGTCAGAAAC

Db	1916	AAGCAGTTTAAACGACGAAGCTGCCAAATGGAATTGGAGAGAGCATATGTCAGAGAAC	1975
Oy	1921	AGGTCACGGGAAGAGCTGGGGGAAAGCTGGGACATGCTAGCTTTTCCGGGACGATGGAA	1980
Db	1976	AGGTCACGGGAAGAGCTGGGGGAAAGTGGGACAGTCAGTCTTACGTTTTCGGGACGATGGAA	2035
Oy	1981	ATCATTTGAGGCTCTCC	1995
Db	2036	ATCATTTGAGGCTCTCC	2050
RESULT 27			
ACC60521			
ID	ACC60521	standard; cDNA; 5111 BP.	
XX	ACC60521;		
XX	19-JUN-2003	(first entry)	
DE	Polynucleotide relating to the invention SEQ ID NO: 41.		
XX	Gene; ss; antiproliferative; hepatotropic; nephrotropic; antiarthritic;		
KW	antiproliferic; cardiant; cytosaratic; gene therapy; liver disease;		
KW	proliferative disorder; renal failure; cardiovascular disorder;		
KW	immunological disorder; arthritis; psoriasis; congenital heart defect;		
KX	congestive heart failure; cancer; human; mouse; yeast; rat; fruitfly.		
OS	Homo sapiens.		
PN	WO200257460-A2.		
XX			
PD	25-JUL-2002.		
XX			
PE	20-DEC-2001; 2001WO-US050459.		
XX			
PR	20-DEC-2000; 2000US-0256868P.		
PR	30-MAR-2001; 2001US-0280186P.		
PR	01-MAY-2001; 2001US-0287735P.		
PR	05-JUN-2001; 2001US-0295848P.		
PR	25-JUN-2001; 2001US-0300465P.		
XX			
PA	(BRIM) BRISTOL-MYERS SQUIBB CO.		
XX			
PI	Jackson DG, Feder J, Nelson T, Mintier G, Ramanathan C, Lee L;		
PI	Stiemers N, Bol D, Schieven G, Finger J, Todderud CG, Bassolino D;		
PI	Kryetsek S, Mcatee P, Suchard S, Ranae D;		
XX			
DR	MPI; 2002-599721/64.		
XX			
PT	P-PSDB; ABR52352.		
XX			
PT	Novel polynucleotides encoding human phosphatase polypeptides useful in		
PT	the prevention or treatment of e.g. proliferative and cardiovascular		
XX	disorders.		
XX			
PS	Claim 1; Fig 5; 801pp; English.		
XX			
CC	The invention relates to a novel isolated nucleic acid comprising a		
CC	polynucleotide having a nucleotide sequence selected from 40		
CC	polynucleotides fully defined in the specification. The polynucleotide of		
CC	the invention has antiproliferative, hepatotropic, nephrotropic,		
CC	antiarthritic, antipsoriatic, cardiant and cytosaratic activity. The		
CC	polypeptide of the invention is useful for preventing, treating or		
CC	ameliorating a medical condition, e.g. a proliferative disorder. They are		
CC	also useful for treating e.g. liver disease, renal failure, immunological		
CC	disorders including arthritic and psoriasis, cardiovascular disorders		
CC	such as congenital heart defects and congestive heart failure, and		
CC	cancer. A method of the invention is useful for diagnosing a pathological		
CC	condition or susceptibility to a condition in a subject. The present		
CC	sequence is used in the exemplification of the invention		
XX			
SO	Sequence 5111 BP; 1413 A; 1112 C; 1232 G; 1354 T; 0 U; 0 Other;		


```
Db 1490 TGTGCCGACTGTCTACCTCAGAGGCAAGAGCAAAAGCCCGTGCAATCCGCCAGGCTG 1549
QY 1081 CCCAGCGGCGCCAGGCTGACCCGCGCTGTTAAGGACAGCCCGCTGGTAACAGGCGCTC 1140
Db 1550 CCCAGCGGCGCCAGGCTGACCCGCGCTGTTAAGGACAGCCCGCTGGTAACAGGCGCTC 1609
QY 1141 AGTGGGCTGCACTGTCCGAGAGAGGCTGGAGAGACAGCAATTAAGCTCAAGCGTCTTC 1200
Db 1610 AGTGGGCTGCACTGTCCGAGAGAGGCTGGAGAGACAGCAATTAAGCTCAAGCGTCTTC 1669
QY 1201 TCTCTGGAATCAATCACTGTTTCATATTACGCGACATGCGAGCATCTTACATGCGTTC 1260
Db 1670 TCTCTGGAATCAATCACTGTTTCATATTACGCGACATGCGAGCATCTTACATGCGTTC 1729
QY 1261 TCTCTCATAGAAAGTGTGGAAATACAAACCTTCCACTACTCTGGAATGGAGCCAC 1320
Db 1730 TCTCTCATAGAAAGTGTGGAAATACAAACCTTCCACTACTCTGGAATGGAGCCAC 1789
QY 1321 AAGCTATGCCAGTCTCCCTGTTACAGAACTATCGAGAGACTCCCGAAACAGTCTCT 1380
Db 1790 AAGCTATGCCAGTCTCCCTGTTACAGAACTATCGAGAGACTCCCGAAACAGTCTCT 1849
QY 1381 GATTAGAGAGAGAGCAGCATCCCGAAGAGCTGACAGCGCCAGGCGTTGACAGCCAG 1440
Db 1850 GATTAGAGAGAGAGCAGCATCCCGAAGAGCTGACAGCGCCAGGCGTTGACAGCCAG 1909
QY 1441 AGCAAGGATTTGGATTGGTCAAGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
Db 1910 AGCAAGGATTTGGATTGGTCAAGAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1969
QY 1501 TCTCCACTGATGTAAGTGGAGAGCGTGGAGAGACAAATTAACAACAGTCTCTTTCGCG 1560
Db 1970 TCTCCACTGATGTAAGTGGAGAGCGTGGAGAGACAAATTAACAACAGTCTCTTTCGCG 2029
QY 1561 CTTTCCACGAGCAGCAGCAGCCTCAAGAGTCTGCTGGGCTTGAAGGCTGGGAC 1620
Db 2030 CTTTCCACGAGCAGCAGCAGCCTCAAGAGTCTGCTGGGCTTGAAGGCTGGGAC 2089
QY 1621 TCGGATATCTTGGCCCCCAGACCTCTACCCCTTCCCTGACAGCAGCTGATTTTGGC 1680
Db 2090 TCGGATATCTTGGCCCCCAGACCTCTACCCCTTCCCTGACAGCAGCTGATTTTGGC 2149
QY 1681 ACGAGTCCCTCACTTACTTGTGCTCAAGCATATACAGAGAGAGAGAGAGAGAGAGAG 1740
Db 2150 ACGAGTCCCTCACTTACTTGTGCTCAAGCATATACAGAGAGAGAGAGAGAGAGAG 2209
QY 1741 GCTTACAGCTGACAGCAGCTGCCACTTGCAGAGACCAAGTCTATTCTGTGGCAGGCGG 1800
Db 2210 GCTTACAGCTGACAGCAGCTGCCACTTGCAGAGACCAAGTCTATTCTGTGGCAGGCGG 2269
QY 1801 CAGAAAGCAGTGAAGAGCTGACTTGCAGGCGAGAGCTGGCATGAAGAGAGAGAGAGAG 1860
Db 2270 CAGAAAGCAGTGAAGAGCTGACTTGCAGGCGAGAGCTGGCATGAAGAGAGAGAGAG 2329
QY 1861 AAGCATTTAAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
Db 2330 AAGCATTTAAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2389
QY 1921 AAGTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
Db 2390 AAGTCAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2449
QY 1981 ATCATTTAGAGTCTCC 1995
Db 2450 ATCATTTAGAGTCTCC 2464
```

```
RESULT 28
AAF30479
XX AAF30479 standard; cDNA, 2118 BP.
AC AAF30479;
```

```
XX
DT 29-MAY-2001 (first entry)
DE Human protein phosphatase and kinase protein-4 cDNA 1234795CB1.
XX
KW Protein phosphatase and kinase protein; PPHK-4; human;
KW gastrointestinal disorder; immune system disorder; neurological disorder;
KW cell proliferative disorder; cancer; diagnosis; therapy; ss.
XX
OS Homo sapiens.
XX
Key Location/Qualifiers
FH 25..2025
FT /*tag= a
FT misc_feature 163..222
FT /*tag= b
FT /*note= "unique fragment"
FT /*tag= c
FT /*note= "unique fragment"
XX
PN WO200120004-A2.
XX
PD 22-MAR-2001.
XX
PF 14-SEP-2000; 2000WO-US025515.
XX
PR 15-SEP-1999; 99US-0154141P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Yue H, Tang YT, Bandman O, Hillman JL, Baughn MR, Azimzai Y,
PI Lu DAM;
XX
DR WPI: 2001-244811/25.
DR P-PSDB; AAB20325.
XX
PT Novel human protein phosphatase and kinase proteins for diagnosis,
PT treatment and prevention of gastrointestinal, immune system, neurological
PT and cell proliferative disorders.
XX
PS Claim 5; Page 97-98; 103pp; English.
XX
CC The present sequence is that of cDNA encoding novel human protein
CC phosphatase and kinase protein PPHK-4 (see AAB20325). The cDNA was
CC initially identified in Incyte Clone ID No. 1234795CB1, from a foetal
CC lung tissue library. Tissues that express PPHK-4 (as a fraction of total
CC tissues expressing PPHK-4) include gastrointestinal (0.385),
CC cardiovascular (0.154), nervous (0.154) and reproductive (0.154).
CC Diseases or conditions associated with tissues expressing PPHK-4 (as a
CC fraction of total tissues expressing PPHK-4) include cancer (0.692),
CC inflammation or trauma (0.308) and cell proliferation (0.231). The
CC encoded protein shows homology to mouse neuronal tyrosine threonine
CC phosphatase 1. The invention provides human PPHK-1 to -11 polypeptides
CC (see AAB20322-32) and polynucleotides (see AAF30476-86). It also provides
CC expression vectors, host cells, antibodies, agonists and antagonists, as
CC well as methods for diagnosing, treating or preventing disorders
CC associated with expression of PPHK, including gastrointestinal
CC disorders, immune system disorders, neurological disorders and cell
CC proliferative disorders, including cancer
XX
SQ Sequence 2118 BP; 550 A; 552 C; 506 G; 510 T; 0 U; 0 Other;
XX
Query Match 79.8%; Score 1593; DB 4; Length 2118;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 3; Gaps 2;
QY 1 ATGGCCCATGAGATGATGGAATGCAATTTTACTGAGAGGTTGAGCTGTGCTGGAA 60
Db 25 ATGGCCCATGAGATGATGGAATGCAATTTTACTGAGAGGTTGAGCTGTGCTGGAA 84
QY 61 AGTGAACGGAAAAAGTGTGCTTAATTGATAGCCGGCCATTTTGGAAATCAATCATCC 120
```

Db 85 AGTGAACGGAAAAAGTGTCTAATGATAGCCGGCATTTGTGNAATACATACATCC 144
QY 121 CAATTTTGAACCAATTAATCACTGCTCAAGTTATGAAGGAAGTTGCAACG 180
Db 145 CAATTTTGAACCAATTAATCACTGCTCAAGTTATGAAGGAAGTTGCAACG 204
QY 181 GACAAAGTGAATTAATCAGAGCTCATCAGCATTCAGCGAAACATTAAGTTGACATTTAT 240
Db 205 GACAAAGTGAATTAATCAGAGCTCATCAGCATTCAGCGAAACATTAAGTTGACATTTAT 264
QY 241 TGCAGTCAGAAAGTTAGTTACGATCAAAAGTCCCAAGATTTGCTCTCTTCA 300
Db 265 TGCAGTCAGAAAGTTAGTTACGATCAAAAGTCCCAAGATTTGCTCTCTCTTCA 324
QY 301 GACTGTTTCTCACTGATCTTCTGGGTAACTGGAAGAGCTTCAACTCTGTTACCTG 360
Db 325 GACTGTTTCTCACTGATCTTCTGGGTAACTGGAAGAGCTTCAACTCTGTTACCTG 384
QY 361 CTTCAGAGTGGGTTTGTGAGTTCTCTGTTGTTTCCCTGGCCCTGTGAAAGGAATCC 420
Db 385 CTTCAGAGTGGGTTTGTGAGTTCTCTGTTGTTTCCCTGGCCCTGTGAAAGGAATCC 444
QY 421 ACTCTAGTCCCTA CTTGATTTCTCAGCTTGTCTTACTGTTGCAACATTTGGCCAC 480
Db 445 ACTCTAGTCCCTA CTTGATTTCTCAGCTTGTCTTACTGTTGCAACATTTGGCCAC 504
QY 481 CGAATTTCTCCCATCTTTATCTTGGCTGCGACGAGATGCTTCAACAGAGCTGATG 540
Db 505 CGAATTTCTCCCATCTTTATCTTGGCTGCGACGAGATGCTTCAACAGAGCTGATG 564
QY 541 CACAGAGTGGGTTTGTGATTTAATGACAGCTATCTGTCCAAAGCTGATCTT 600
Db 565 CACAGAGTGGGTTTGTGATTTAATGACAGCTATCTGTCCAAAGCTGATCTT 624
QY 601 ATCCCGAGTCTCAATTTCTGCGGTGCTGTGAATGACAGCTTTGTGAAATTTTG 660
Db 625 ATCCCGAGTCTCAATTTCTGCGGTGCTGTGAATGACAGCTTTGTGAAATTTTG 684
QY 661 CCGTGGTTGACAAATCAGTGAATTTTCAATTTGAAGAAAGCAAGCTTCAATTTGAT 720
Db 685 CCGTGGTTGACAAATCAGTGAATTTTCAATTTGAAGAAAGCAAGCTTCAATTTGAT 744
QY 721 CTAGTGAATGTTAGTGGGATCTCCGCTCGGCCCATGCTATGCTATCATCATG 780
Db 745 CTAGTGAATGTTAGTGGGATCTCCGCTCGGCCCATGCTATGCTATCATCATG 804
QY 781 AAGAGATGACATGTTCTTGAATGAAGCTTACAGATTTGTGAAGAAAGAAAGCACTACT 840
Db 805 AAGAGATGACATGTTCTTGAATGAAGCTTACAGATTTGTGAAGAAAGAAAGCACTACT 864
QY 841 ATATCTCCAACTTCAATTTTCTGGGCAACTCTGATCTATGAAGAAAGTTAAAGAC 900
Db 865 ATATCTCCAACTTCAATTTTCTGGGCAACTCTGATCTATGAAGAAAGTTAAAGAC 924
QY 901 CAGACTGAGACATCAGGGCCAAAGAGCAACTCAAGCTGTGACCTGGAAGGCAAAAT 960
Db 925 CAGACTGAGACATCAGGGCCAAAGAGCAACTCAAGCTGTGACCTGGAAGGCAAAAT 984
QY 961 GAACTGTCCCTGCTCTCAGAGGGTGA CAGAAAAAGCAGAGCCCTCACTCAACCC 1020
Db 985 GAACTGTCCCTGCTCTCAGAGGGTGA CAGAAAAAGCAGAGCCCTCACTCAACCC 1044
QY 1021 TGTGCGGACTCTGCTACATCAGAGGAGCAGAGCAAAAGCCGTGATCCCGCAGCGT 1080
Db 1045 TGTGCGGACTCTGCTACATCAGAGGAGCAGAGCAAAAGCCGTGATCCCGCAGCGT 1104
QY 1081 CCGAGGCTGCGAGCGTGAAGCGTGTGTTAGAGGAGAGCCGCTGATCAGGGGCTT 1140
Db 1105 CCGAGGCTGCGAGCGTGAAGCGTGTGTTAGAGGAGAGCCGCTGATCAGGGGCTT 1164
QY 1141 AGTGGCTGCACTGTCTCGAGAGAGGCTGGAAGAGAGCAATTAAGCTCAAGCGTTCTTC 1200
Db 1165 AGTGGCTGCACTGTCTCGAGAGAGGCTGGAAGAGAGCAATTAAGCTCAAGCGTTCTTC 1224

QY 1201 TCTCTGATATGAATCAGTTTCAATTCAGGACAGATGGAGCATCTTACATGGCTTC 1260
Db 1225 TCTCTGATATGAATCAGTTTCAATTCAGGACAGATGGAGCATCTTACATGGCTTC 1284
QY 1261 TCTCATCAGAAAGTCTTTGGAACTACTA CAAACCTTTCATCTGTGAATGGACCAAC 1320
Db 1285 TCTCATCAGAAAGTCTTTGGAACTACTA CAAACCTTTCATCTGTGAATGGACCAAC 1344
QY 1321 AAGCTATGCGAGTTCTCCCTGTTACAGAACTATCGAGAGACTCCGAAACAGT--C 1378
Db 1345 AAGCTATGCGAGTTCTCCCTGTTACAGAACTATCGAGAGACTCCGAAACAGT--C 1404
QY 1379 CTGATTAAGGAGAGCAGCATCCCAAGAGTGAAGCCGCGAGGCTTGAAGAGCC 1438
Db 1405 CTGATTAAGGAGAGCAGCATCCCAAGAGTGAAGCCGCGAGGCTTGAAGAGCC 1464
QY 1439 AGAGCAAGGATTCATTCGCTCAGAACAGCAGCAGTGGCAACCGCCAGAGGTCCTTT 1498
Db 1465 AGAGCAAGGATTCATTCGCTCAGAACAGCAGCAGTGGCAACCGCCAGAGGTCCTTT 1524
QY 1499 TATCTCACTGATGGAAGTGGAGCTGGAAGCAATTAACAACAGCTTCTTTTG 1558
Db 1525 TATCTCACTGATGGAAGTGGAGCTGGAAGCAATTAACAACAGCTTCTTTTG 1584
QY 1559 GCCTTCCACAGCGCAGCAGCCTTCAAGAGTCTGCTGGCCCTGGG--CCTTAAAGGCTGG 1617
Db 1585 GCCTTCCACAGCGCAGCAGCCTTCAAGAGTCTGCTGGCCCTGGG--CCTTAAAGGCTGG 1644
QY 1618 CACTCGGATATCTTGGCCCCCAGACCTTACCCCTTCCCTGACAGAGCTGTATTTT 1677
Db 1645 CACTCGGATATCTTGGCCCCCAGACCTTACCCCTTCCCTGACAGAGCTGTATTTT 1704
QY 1678 GCCACAGATCTTACACTTTCTACTGCTTCAAGCCATCTAGGAGGAGTGGCACTTAC 1737
Db 1705 GCCACAGATCTTACACTTTCTACTGCTTCAAGCCATCTAGGAGGAGTGGCACTTAC 1764
QY 1738 TCTGCTTACAGCTGAGCGAGCTGCCACTTGGGAGCAAGTCTTCTGTGGCGAG 1797
Db 1765 TCTGCTTACAGCTGAGCGAGCTGCCACTTGGGAGCAAGTCTTCTGTGGCGAG 1824
QY 1798 CCGCAGAAAGCAAGTACAGAGCTGCTGCGCGAGCTGCGATGAAGAGAGCCCTTT 1857
Db 1825 CCGCAGAAAGCAAGTACAGAGCTGCTGCGCGAGCTGCGATGAAGAGAGCCCTTT 1884
QY 1858 GAAAGCACTTTAAACGCAAGAGCTGCCAAATGGAATTTGAGAGAGCATCATGTCAAG 1917
Db 1885 GAAAGCACTTTAAACGCAAGAGCTGCCAAATGGAATTTGAGAGAGCATCATGTCAAG 1944
QY 1918 AACAGTCAAGGGAAGAGCTGGGGAAGTGGGCAAGTCAATCTTTTCCGGCAGCATG 1977
Db 1945 AACAGTCAAGGGAAGAGCTGGGGAAGTGGGCAAGTCAATCTTTTCCGGCAGCATG 2004
QY 1978 GAAATCAATGAGTCTCC 1995
Db 2005 GAAATCAATGAGTCTCC 2022

Search completed: June 22, 2004, 10:38:26
Job time : 1150 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2004, 10:19:15 ; Search time 1174 Seconds

(without alignments)
7784.730 Million cell updates/sec

Title: US-10-029-345a-108_COPY_538_2532

Perfect score: 1995
Sequence: 1 atggcccatgagatgattgttg.....tggaaatcatgaggtctcc 1995

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3017426 seqs, 2290544650 residues

Word size : 1419

Total number of hits satisfying chosen parameters: 12

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 75 summaries

Database :

Published Applications NA:
1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq:*
3: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq:*
4: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq:*
5: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq:*
6: /cgn2_6/prodata/2/pubpna/PCUS_PUBCOMB.seq:*
7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/prodata/2/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/prodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/prodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/prodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/prodata/2/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1893	94.9	1998	9 US-09-816-494-3	Sequence 3, Appl
2	1893	94.9	1998	16 US-10-377-072-27	Sequence 27, Appl
3	1893	94.9	3059	17 US-10-257-026-1	Sequence 1, Appl
4	1893	94.9	3496	9 US-09-964-277-1	Sequence 1, Appl
5	1893	94.9	3544	9 US-09-816-494-1	Sequence 1, Appl
6	1893	94.9	3544	16 US-10-377-072-25	Sequence 25, Appl
7	1893	94.9	3625	13 US-10-425-114-26234	Sequence 26234, A
8	1893	94.9	3766	13 US-10-343-357-17	Sequence 17, Appl
9	1893	94.9	4790	17 US-10-648-593-115	Sequence 115, App
10	1849	92.7	2966	13 US-10-296-115-520	Sequence 520, App
11	1842	92.3	2732	13 US-10-168-506-2	Sequence 2, Appl
12	1791	89.8	2102	16 US-10-094-749-673	Sequence 673, App

ALIGNMENTS

RESULT 1
US-09-816-494-3
; Sequence 3, Application US/09816494
; Patent No. US20020034807A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel A.
; TITLE OF INVENTION: 38692 AND 21117, NOVEL DUAL SPECIFICITY
; FILE REFERENCE: 10448-030002
; CURRENT APPLICATION NUMBER: US/09/816,494
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 60/191,858
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1998
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-816-494-3

Query Match 94.9%; Score 1893; DB 9; Length 1998;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY	1	ATGGCCCATGAGATGATGGAACCTCAATTTGTTACTGAGAGTTGGCTGCTGGA	60
DB	1	ATGGCCCATGAGATGATGGAACCTCAATTTGTTACTGAGAGTTGGCTGCTGGA	60
QY	61	AGTGAACCGGAAAAAGTCTGCTTAATTGATAGCCGCCATTTGTGAATACATAC	120
DB	61	AGTGAACCGGAAAAAGTCTGCTTAATTGATAGCCGCCATTTGTGAATACATAC	120
QY	121	CACATTTTGAAGCCATTATATCACTGCTCCAGCTTATGAGGAAAGTTGCAAC	180
DB	121	CACATTTTGAAGCCATTATATCACTGCTCCAGCTTATGAGGAAAGTTGCAAC	180
QY	181	GACAAAGTTTATTAACAGAGCTCATCAGATTCAGCGAACAATAGTTGACAT	240
DB	181	GACAAAGTTTATTAACAGAGCTCATCAGATTCAGCGAACAATAGTTGACAT	240
QY	241	TGACGTCAGAAAGTTTATTAACATCAAGTCCCAAGATTTGCTCTCTTCA	300
DB	241	TGACGTCAGAAAGTTTATTAACATCAAGTCCCAAGATTTGCTCTCTTCA	300
QY	301	GACTGTTTCTACCTGACTCTGGGTAAACTGGAGAAGCTTCAACTCTGTCA	360
DB	301	GACTGTTTCTACCTGACTCTGGGTAAACTGGAGAAGCTTCAACTCTGTCA	360
QY	361	CTTGACAGTGGTTTCTGAGATTTCTGTTTCTGCTGCTGCTGTAAGAAATCC	420
DB	361	CTTGACAGTGGTTTCTGAGATTTCTGTTTCTGCTGCTGCTGTAAGAAATCC	420
QY	421	ACTTATGTCCTACCTGATTTCTGAGCTTGTCTTACCTGTTGCCAATTTGG	480
DB	421	ACTTATGTCCTACCTGATTTCTGAGCTTGTCTTACCTGTTGCCAATTTGG	480
QY	481	CGAATCTTCCCAATTTTATCTGGCTGCGAGCGAGATGCTTCAACAGAGCTGA	540
DB	481	CGAATCTTCCCAATTTTATCTGGCTGCGAGCGAGATGCTTCAACAGAGCTGA	540
QY	541	CAGCAAAATGGAGTTGTTATGTTAAATGCGAGCTATACCTGTCAAAGCTGAC	600
DB	541	CAGCAAAATGGAGTTGTTATGTTAAATGCGAGCTATACCTGTCAAAGCTGAC	600
QY	601	ATCCCGAGTCTCATTTCTGCGTGTGCTGCTGATGACAGCTTTTGTGAGAAAT	660
DB	601	ATCCCGAGTCTCATTTCTGCGTGTGCTGCTGATGACAGCTTTTGTGAGAAAT	660
QY	661	CGGTGTTGGAACAATCAGATTTTCAATTGAGAAAGCAAAAGCTTCAATGATG	720

Db 661 CCGTGGTGGACAAATCAGTATGATTTCTTGGAGAAAGCAAAAGCCCTCCAAATGAGATGTT 720
Qy 721 CTAGTGCATCTTTTAGCTGGGATCTCCGCTCCGCCACCATCGCTATGCGCTACATCATG 780
Db 721 CTAGTGCATCTTTTAGCTGGGATCTCCGCTCCGCCACCATCGCTATGCGCTACATCATG 780
Qy 781 AAGAGGATGACATGCTTTAGATGAAGCTTCAATTTTGGAAAGAAAGAACCTACT 840
Db 781 AAGAGGATGACATGCTTTAGATGAAGCTTCAATTTTGGAAAGAAAGAACCTACT 840
Qy 841 AATATCTCCAACTTCATTTTCTGGCCAACTCCGCTGACTATGAGAAAGATTTAAGAAC 900
Db 841 AATATCTCCAACTTCATTTTCTGGCCAACTCCGCTGACTATGAGAAAGATTTAAGAAC 900
Qy 901 CAGACTGGAGATCGAGGCGCAAGAGCAAACTCAAGCTGCTGCACTGAGAAAGCCAAAT 960
Db 901 CAGACTGGAGATCGAGGCGCAAGAGCAAACTCAAGCTGCTGCACTGAGAAAGCCAAAT 960
Qy 961 GAACCTGTCCCTGCTGCTCAGAGGGGTGAGCAAGAAAGCGAGCGCCCTCAGTCCACCC 1020
Db 961 GAACCTGTCCCTGCTGCTCAGAGGGGTGAGCAAGAAAGCGAGCGCCCTCAGTCCACCC 1020
Qy 1021 TGTGCGGACTCTGCTACTCTCAGAGGAGAGCAAGAGCGCCGCTGATCCCGCAGCGTG 1080
Db 1021 TGTGCGGACTCTGCTACTCTCAGAGGAGAGCAAGAGCGCCGCTGATCCCGCAGCGTG 1080
Qy 1081 CCCAGCGTCCGAGGCTGAGCGTCCGCTTTAGAGAGAGCGCGGCTGATCAGAGGCTC 1140
Db 1081 CCCAGCGTCCGAGGCTGAGCGTCCGCTTTAGAGAGAGCGCGGCTGATCAGAGGCTC 1140
Qy 1141 AGTGGGCTGCACTGTCCGAGAGAGCGCTGAAAGACAGCAATTAAGCTCAAGCCTTCTTC 1200
Db 1141 AGTGGGCTGCACTGTCCGAGAGAGCGCTGAAAGACAGCAATTAAGCTCAAGCCTTCTTC 1200
Qy 1201 TCTGTGATATCAATAGTTCATTTATTCAGCCAGCATGCGACATCTTACATGCTTC 1260
Db 1201 TCTGTGATATCAATAGTTCATTTATTCAGCCAGCATGCGACATCTTACATGCTTC 1260
Qy 1261 TCTGTGATCAGAGATGCTTTGGAATCTAATAAATCTTCACTCTGAGTGGAGCAAC 1320
Db 1261 TCTGTGATCAGAGATGCTTTGGAATCTAATAAATCTTCACTCTGAGTGGAGCAAC 1320
Qy 1321 AAGCTATGCAAGTTCCTCTGTTCAAGAACTATCGAGCACTCCGAAACCAAGTCTCT 1380
Db 1321 AAGCTATGCAAGTTCCTCTGTTCAAGAACTATCGAGCACTCCGAAACCAAGTCTCT 1380
Qy 1381 GATTAAGAGAGAGCCAGCATCCCAAGAAAGCTGCAAGCGCGCTTCAAGAGCCAG 1440
Db 1381 GATTAAGAGAGAGCCAGCATCCCAAGAAAGCTGCAAGCGCGCTTCAAGAGCCAG 1440
Qy 1441 AGCAAGCGATTCGATTCGATCAGAACCAAGCAAGTGGCAGCGCCAGAGTCCCTTTTA 1500
Db 1441 AGCAAGCGATTCGATTCGATCAGAACCAAGCAAGTGGCAGCGCCAGAGTCCCTTTTA 1500
Qy 1501 TCTCCACTGATGAGAGTGGAGCGTGGAGCAATTTACACAACAGCTCTCTTTGGGC 1560
Db 1501 TCTCCACTGATGAGAGTGGAGCGTGGAGCAATTTACACAACAGCTCTCTTTGGGC 1560
Qy 1561 CTTTCCACAGCAGCAGGAGCACTCAAGAAAGTGTGCTGCGCTTGAAGGCTGGGAC 1620
Db 1561 CTTTCCACAGCAGCAGGAGCACTCAAGAAAGTGTGCTGCGCTTGAAGGCTGGGAC 1620
Qy 1621 TGGGATATCTGGGCCCCCAGACTCTTACCTTCCCTGACCAAGAGCTGTATTGGCC 1680
Db 1621 TGGGATATCTGGGCCCCCAGACTCTTACCTTCCCTGACCAAGAGCTGTATTGGCC 1680
Qy 1681 ACAGAGTCTTCACTTCTAGCTGCTCAGGCACTTACAGAGAGAGTCCAGTTACTCT 1740
Db 1681 ACAGAGTCTTCACTTCTAGCTGCTCAGGCACTTACAGAGAGAGTCCAGTTACTCT 1740
Qy 1741 GCTTACAGTGGAGCAGCTGCCCATTTGCGAGAGCAAGTCTATTCTGTGCGAGCGG 1800
Db 1741 GCTTACAGTGGAGCAGCTGCCCATTTGCGAGAGCAAGTCTATTCTGTGCGAGCGG 1800

Db 1741 GCTTACAGTGGAGCAGCTGCCCATTTGCGAGAGCAAGTCTATTCTGTGCGAGCGG 1800
Qy 1801 CAGAGCGAAGTACAGAGCTGACTGCGCGGAGCTGGCATGAAAGAGCCCTTTGAA 1860
Db 1801 CAGAGCGAAGTACAGAGCTGACTGCGCGGAGCTGGCATGAAAGAGCCCTTTGAA 1860
Qy 1861 AAGCAGTTTAAAGCAGAGCTGCCAAATGAAATTTGGAGAGGATCATGTCAGAGAAC 1920
Db 1861 AAGCAGTTTAAAGCAGAGCTGCCAAATGAAATTTGGAGAGGATCATGTCAGAGAAC 1920
Qy 1921 AGGTACCGGAAAGAGTGGGAAAGTGGGAGTCACTTACCTTTTGGCGCAGATGAA 1980
Db 1921 AGGTACCGGAAAGAGTGGGAAAGTGGGAGTCACTTACCTTTTGGCGCAGATGAA 1980
Qy 1981 ATCATTGAGGTCTCC 1995
Db 1981 ATCATTGAGGTCTCC 1995

541-1945P 1455

RESULT 2
US-10-377-072-27
; Sequence 27, Application US/10377072
; Publication No. US20040009501A1
GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals Inc.
; APPLICANT: Curtis, Rory A.J.
; APPLICANT: Logan, Thomas Joseph
; APPLICANT: Glucksmann, Maria A.
; APPLICANT: Meyers, Rachel E.
; APPLICANT: Williamson, Mark J.
; APPLICANT: Rudolph-Owen, Laura A.
; APPLICANT: Tsai, Fong-Ying
; TITLE OF INVENTION: NOVEL 25869, 25934, 26335, 50365, 21117,
; TITLE OF INVENTION: 38692, 46508, 16816, 16839, 49937, 49931 AND 49933 MOLECULES
; FILE REFERENCE: MP103-0180NM19
; CURRENT APPLICATION NUMBER: US/10/377, 072
; PRIOR FILING DATE: 2003-02-27
; PRIOR APPLICATION NUMBER: US 09/895,860
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/215,370
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: US 09/723,806
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: US 60/187,455
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: US 09/843,297
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: US 60/199,801
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: US 09/861,801
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: US 60/205,508
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: US 09/816,494
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/815,419
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 1998
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1998)
US-10-377-072-27

Query Match 94.9%; Score 1893; DB 16; Length 1998;
Best Local Similarity 99.9%; Pred. No. 0;

FILE REFERENCE: DUSP10X.DMS
CURRENT APPLICATION NUMBER: US/10/257,026
CURRENT FILING DATE: 2003-11-07
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 3059
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (127)..(2121)
US-10-257-026-1

Query Match 94.9%; Score 1893; DB 17; Length 3059;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCCCATGAGTATATGGAATCTCAATTGTACTGAGAGTTGGCTCTGCTGAA 60
DB 127 ATGGCCCATGAGTATATGGAATCTCAATTGTACTGAGAGTTGGCTCTGCTGAA 186
QY 61 AGTGAACGGAAGAAAGTCTGCTAATGATAGCCGCCATTTGTGAATCAATACATCC 120
DB 187 AGTGAACGGAAGAAAGTCTGCTAATGATAGCCGCCATTTGTGAATCAATACATCC 246
QY 121 CACATTTTGAAGCCATTATATATCACTGCTCCAAAGCTTATGAAGCGAGTTGCAACG 180
DB 247 CACATTTTGAAGCCATTATATATCACTGCTCCAAAGCTTATGAAGCGAGTTGCAACG 306
QY 181 GACAAAGTGTATATTAAGAGCTATCCAGATTCAGCGAAGCAATTAAGTTGACATTTGAT 240
DB 307 GACAAAGTGTATATTAAGAGCTATCCAGATTCAGCGAAGCAATTAAGTTGACATTTGAT 366
QY 241 TGCATGAGAGGTTGATGATTAAGATCAAGCTCCCAAGATGTTGCTCTCTCTTCA 300
DB 367 TGCATGAGAGGTTGATGATTAAGATCAAGCTCCCAAGATGTTGCTCTCTCTTCA 426
QY 301 GACTGTTTCTCACTGTACTTCTGGGTAACTGAGAGAGCTTCAACTCTGTTCACTG 360
DB 427 GACTGTTTCTCACTGTACTTCTGGGTAACTGAGAGAGCTTCAACTCTGTTCACTG 486
QY 361 CTGCAAGTGGGTTGGTGGTCTCTCGTTGTTTCCCTGGCTCTGTGAAGGAAATCC 420
DB 487 CTGCAAGTGGGTTGGTGGTCTCTCGTTGTTTCCCTGGCTCTGTGAAGGAAATCC 546
QY 421 ACTCTAGTCCCTACCTGCAATTTCTAGCCCTTGTCTTACCTGTTGCGCAAC 480
DB 547 ACTCTAGTCCCTACCTGCAATTTCTAGCCCTTGTCTTACCTGTTGCGCAAC 606
QY 481 CGAATTCTTCCCAATCTTTATCTGGCTGCCAGCGAGATGCTTCAACAGAGGCTGATA 540
DB 607 CGAATTCTTCCCAATCTTTATCTGGCTGCCAGCGAGATGCTTCAACAGAGGCTGATA 666
QY 541 CAGCAGAAATGGGATGTTATGTTAAATGCAAGTAACTGTGCCAAAGCTTCACTTT 600
DB 667 CAGCAGAAATGGGATGTTATGTTAAATGCAAGTAACTGTGCCAAAGCTTCACTTT 726
QY 601 ATCCCGAATCTTCAATTTCTGCGTGTGCTGTGAATGACAGCTTTTGAAGAAATTTTG 660
DB 727 ATCCCGAATCTTCAATTTCTGCGTGTGCTGTGAATGACAGCTTTTGAAGAAATTTTG 786
QY 661 CCGTGTGGGACAAATACGTATATTTCAATGAGAAAGCAAAAGCTTCAATGATGTGT 720
DB 787 CCGTGTGGGACAAATACGTATATTTCAATGAGAAAGCAAAAGCTTCAATGATGTGT 846
QY 721 CTAGTGAATGTTAGTGGGATCTCCGCTCCGCAACATGCGTATGCTTACATATG 780
DB 847 CTAGTGAATGTTAGTGGGATCTCCGCTCCGCAACATGCGTATGCTTACATATG 906
QY 781 AAGAGAGTGAATGCTTTTGAATGAGCTTACAGATTTGTGAAGGAAAGAAAGCTTACT 840
DB 907 AAGAGAGTGAATGCTTTTGAATGAGCTTACAGATTTGTGAAGGAAAGAAAGCTTACT 966

QY 841 ATATCTCAAACTTCAATTTTCTGGGCCAACTCTGGACTATAGAGATTAAGAAC 900
DB 967 ATATCTCAAACTTCAATTTTCTGGGCCAACTCTGGACTATAGAGATTAAGAAC 1026
QY 901 CAGACTGAGCATCAGGGCCAAAGCAAACTCAAGCTCTGACCTGAGAGGCAAT 960
DB 1027 CAGACTGAGCATCAGGGCCAAAGCAAACTCAAGCTCTGACCTGAGAGGCAAT 1086
QY 961 GAACTGTCCCTGCTGTCTCAGAGGTTGAGCAAAAGCGAGAGCCCTCTCACTCACCC 1020
DB 1087 GAACTGTCCCTGCTGTCTCAGAGGTTGAGCAAAAGCGAGAGCCCTCTCACTCACCC 1146
QY 1021 TGTGCCAATCTGTACTCTCAGAGGCAAGAGCAAAAGCCCGTACATCCCGCAGGTG 1080
DB 1147 TGTGCCAATCTGTACTCTCAGAGGCAAGAGCAAAAGCCCGTACATCCCGCAGGTG 1206
QY 1081 CCAGCGTCCCGAGGTGACAGCCGTCTGTTAAGAGACAGCCCGTGTGACAGGCGCTC 1140
DB 1207 CCAGCGTCCCGAGGTGACAGCCGTCTGTTAAGAGACAGCCCGTGTGACAGGCGCTC 1266
QY 1141 AGTGGCTGCACTGTCCGAGACAGGCTGGAAGACAAATAAGCTCAAGCTTCTTC 1200
DB 1267 AGTGGCTGCACTGTCCGAGACAGGCTGGAAGACAAATAAGCTCAAGCTTCTTC 1326
QY 1201 TCTGTGATATCAATCACTGTTTCAATTCAGCCAGATGCGAGCATCTTACATGCTTC 1260
DB 1327 TCTGTGATATCAATCACTGTTTCAATTCAGCCAGATGCGAGCATCTTACATGCTTC 1386
QY 1261 TCTCATAGAGAGATGCTTGGAACTACTAACAACCTTCCACTACTGTGATGAGACCAAC 1320
DB 1387 TCTCATAGAGAGATGCTTGGAACTACTAACAACCTTCCACTACTGTGATGAGACCAAC 1446
QY 1321 AAGCTATGCCAGTTCTCCCTGTTCAAGAACTATCGAGAGACATCCCGAAACAGGCTT 1380
DB 1447 AAGCTATGCCAGTTCTCCCTGTTCAAGAACTATCGAGAGACATCCCGAAACAGGCTT 1506
QY 1381 GATTAAGAGAGAGGCAAGCTATCCCAAGAGCTGACACGCGCAGGCTTCAAGACAG 1440
DB 1507 GATTAAGAGAGAGGCAAGCTATCCCAAGAGCTGACACGCGCAGGCTTCAAGACAG 1566
QY 1441 AGCAAGGATTTGCAATTCGGTCAAGACAGAGAGTGGGACCGCCAGAGGTCCTTTTA 1500
DB 1567 AGCAAGGATTTGCAATTCGGTCAAGACAGAGAGTGGGACCGCCAGAGGTCCTTTTA 1626
QY 1501 TCTCACTGATGAAAGTGGAGCTGTGAAGACAAATTACACACAGCTTCTTTTCGCG 1560
DB 1627 TCTCACTGATGAAAGTGGAGCTGTGAAGACAAATTACACACAGCTTCTTTTCGCG 1686
QY 1561 CTTTCCACAGCAGAGAGACCTTCAAGAGTGTGCTGGCTTGAAGGCTGGCAC 1620
DB 1687 CTTTCCACAGCAGAGAGACCTTCAAGAGTGTGCTGGCTTGAAGGCTGGCAC 1746
QY 1621 TCGGATATCTTGGCCCCCAGACCTTACCCCTTCTGACAGAGGCTGATTTTGGC 1680
DB 1747 TCGGATATCTTGGCCCCCAGACCTTACCCCTTCTGACAGAGGCTGATTTTGGC 1806
QY 1681 ACAGATCTCTCACTTTACTGTGCTTCAAGCCATCTACAGAGGAGTCCAGTTACTCT 1740
DB 1807 ACAGATCTCTCACTTTACTGTGCTTCAAGCCATCTACAGAGGAGTCCAGTTACTCT 1866
QY 1741 GCCTACAGCTGACAGCAGCTGCGCACTTGGCGAGACCAAGTCTATCTGTGCGCAGGCGG 1800
DB 1867 GCCTACAGCTGACAGCAGCTGCGCACTTGGCGAGACCAAGTCTATCTGTGCGCAGGCGG 1926
QY 1801 CAGAAAGCAAGTGAAGAGCTGACTCGCGCGAGCTGGCAATGAAGAGCCCTTTGAA 1860
DB 1927 CAGAAAGCAAGTGAAGAGCTGACTCGCGCGAGCTGGCAATGAAGAGCCCTTTGAA 1986
QY 1861 AAGCAGTTTAAACGAGAAAGCTGCGCAATGGAATTTGAGAGAGCATCATCAGAGAAC 1920
DB 1987 AAGCAGTTTAAACGAGAAAGCTGCGCAATGGAATTTGAGAGAGCATCATCAGAGAAC 2046

QY 1921 AGGTCAAGGAGAGAGTGGGAGAGTGGAGCTAGTCTAGTCTTTGGGGCAGCATGAA 1980
| | | | |
Db 2047 AGGTCAAGGAGAGAGTGGGAGAGTGGAGCTAGTCTAGTCTTTGGGGCAGCATGAA 2106
| | | | |
QY 1981 ATCATTGAGGTCTCC 1995
| | | | |
Db 2107 ATCATTGAGGTCTCC 2121
| | | | |

RESULT 4
US-09-964-277-1
; Sequence 1, Application US/09964277
; Patent No. US20020137170A1
; GENERAL INFORMATION:
; APPLICANT: Lucbe, Ralf M.
; APPLICANT: Wei, Bo
; TITLE OF INVENTION: DSP-16 DUAL-SPECIFICITY PHOSPHATASE
; FILE REFERENCE: 200125.434
; CURRENT APPLICATION NUMBER: US/09/964,277
; CURRENT FILING DATE: 2001-09-25
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3496
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-964-277-1

Query Match 94.9%; Score 1893; DB 9; Length 3496;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGCCCCATAGATGATGATGGAATCTCAATGTTTACTGAGAGGTTGGGCTCTGCTGGA 60
| | | | |
Db 562 ATGCCCCATAGATGATGATGGAATCTCAATGTTTACTGAGAGGTTGGGCTCTGCTGGA 621
| | | | |
QY 61 AGTGAACGGAAGAGTGTCTAATGATGAGCGGCAATTTGGAATACATATGATCC 120
| | | | |
Db 622 AGTGAACGGAAGAGTGTCTAATGATGAGCGGCAATTTGGAATACATATGATCC 661
| | | | |
QY 121 CACATTTTGGAGCCATTAATATCACTGCTCCAGGTTATGAAGGAGGTTGCAACAG 180
| | | | |
Db 682 CACATTTTGGAGCCATTAATATCACTGCTCCAGGTTATGAAGGAGGTTGCAACAG 741
| | | | |
QY 181 GACAAAGTGTATTAACAGAGCTCATCCAGCATTCAGGAAACATTAAGTTGATGAT 240
| | | | |
Db 742 GACAAAGTGTATTAACAGAGCTCATCCAGCATTCAGGAAACATTAAGTTGATGAT 801
| | | | |
QY 241 TGCAGTCAGAGGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
| | | | |
Db 802 TGCAGTCAGAGGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 861
| | | | |
QY 301 GACTGTTTCTCACTGATCTTGGGTAACCTGGAGAGAGTTCACATCTGTTCACTG 360
| | | | |
Db 862 GACTGTTTCTCACTGATCTTGGGTAACCTGGAGAGAGTTCACATCTGTTCACTG 921
| | | | |
QY 361 CTTCGAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 420
| | | | |
Db 922 CTTCGAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 981
| | | | |
QY 421 ACTCTAGTCCCTCACTGATCTTCAAGCCTTGGTTCAGTGGTGGTGGTGGTGGTGGTGG 480
| | | | |
Db 982 ACTCTAGTCCCTCACTGATCTTCAAGCCTTGGTTCAGTGGTGGTGGTGGTGGTGGTGG 1041
| | | | |
QY 481 CGAATTTCTCCATCTTTATCTTGGCTGCGCAGAGATGCTCTCAACAGAGCTGATA 540
| | | | |
Db 1042 CGAATTTCTCCATCTTTATCTTGGCTGCGCAGAGATGCTCTCAACAGAGCTGATA 1101
| | | | |
QY 541 CAGCAGATGGAGTGGTATGTTGTTAATGCAAGCTATACCTGTCCAAAAGCTGACTT 600
| | | | |
Db 1102 CAGCAGATGGAGTGGTATGTTGTTAATGCAAGCTATACCTGTCCAAAAGCTGACTT 1161
| | | | |
QY 601 ATCCCGAGTCTCATTTCTGCGGTGGCTGTGATGACAGCTTTTGTGAGAAATTTTG 660
| | | | |

Db 1162 ATCCCGAGTCTCATTTCTGCGGTGGCTGTGATGACAGCTTTTGTGAGAAATTTTG 1221
| | | | |
QY 661 CCGTGTGGAGCAATCAGTATGATTTTCAATTGAGAAAGCAAAAGCTTCAATGATGTT 720
| | | | |
Db 1222 CCGTGTGGAGCAATCAGTATGATTTTCAATTGAGAAAGCAAAAGCTTCAATGATGTT 1281
| | | | |
QY 721 CTAGTGCACCTGTTTACTGAGTCTTCCGCTCCGCAATGCTATGCTATGATGATGAT 780
| | | | |
Db 1282 CTAGTGCACCTGTTTACTGAGTCTTCCGCTCCGCAATGCTATGCTATGATGATGAT 1341
| | | | |
QY 781 AAGAGATGAGCAATGCTTTTATGAGCTTACAGATTTTGTGAGAAAGCAAAAGCTTACT 840
| | | | |
Db 1342 AAGAGATGAGCAATGCTTTTATGAGCTTACAGATTTTGTGAGAAAGCAAAAGCTTACT 1401
| | | | |
QY 841 ATATCTCCAACTTCAATTTTCTGGGCAATCTCTGACTATGAGAAAGATTAAGAAC 900
| | | | |
Db 1402 ATATCTCCAACTTCAATTTTCTGGGCAATCTCTGACTATGAGAAAGATTAAGAAC 1461
| | | | |
QY 901 CAGACTGAGCATCAGGCGCAAGCAAACTCAAGCTGCTGCACTGAGAGAGCCCAAT 960
| | | | |
Db 1462 CAGACTGAGCATCAGGCGCAAGCAAACTCAAGCTGCTGCACTGAGAGAGCCCAAT 1521
| | | | |
QY 961 GAACTGTCTCTGCTGTCTCAGAGGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
| | | | |
Db 1522 GAACTGTCTCTGCTGTCTCAGAGGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1581
| | | | |
QY 1021 TGTGCGACTGTGTAATCTCAGAGGAGAGAGCAAAAGCCCTGATCCCGCAGCGTG 1080
| | | | |
Db 1582 TGTGCGACTGTGTAATCTCAGAGGAGAGAGCAAAAGCCCTGATCCCGCAGCGTG 1641
| | | | |
QY 1081 CCCAGGTGCGCAGAGCTGAGAGCTGCTGTTTGAAGAGAGAGAGAGAGAGAGAGAGAG 1140
| | | | |
Db 1642 CCCAGGTGCGCAGAGCTGAGAGCTGCTGTTTGAAGAGAGAGAGAGAGAGAGAGAGAG 1701
| | | | |
QY 1141 AGTGGCTGCACTGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
| | | | |
Db 1702 AGTGGCTGCACTGCTCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1761
| | | | |
QY 1201 TCTCTGATATCAATCAGTTTATATTCAGCAGCAGTGGAGCATCTTACATGCTTC 1260
| | | | |
Db 1762 TCTCTGATATCAATCAGTTTATATTCAGCAGCAGTGGAGCATCTTACATGCTTC 1821
| | | | |
QY 1261 TCTCTATGAGAGATGCTTTGGAATCTCAAACTTCCATCTTGTGATGGAGCAAC 1320
| | | | |
Db 1822 TCTCTATGAGAGATGCTTTGGAATCTCAAACTTCCATCTTGTGATGGAGCAAC 1881
| | | | |
QY 1321 AAGCTATGCAATGCTCCCTGTTTCAAGAACTATCGAGAGAGAGAGAGAGAGAGAGAG 1380
| | | | |
Db 1882 AAGCTATGCAATGCTCCCTGTTTCAAGAACTATCGAGAGAGAGAGAGAGAGAGAGAG 1941
| | | | |
QY 1381 GATTAAG 1440
| | | | |
Db 1942 GATTAAG 2001
| | | | |
QY 1441 AGCAGAGATGATGATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1500
| | | | |
Db 2002 AGCAGAGATGATGATGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2061
| | | | |
QY 1501 TCTCCAGTCACTGAG 1560
| | | | |
Db 2062 TCTCCAGTCACTGAG 2121
| | | | |
QY 1561 CTTTCACAG 1620
| | | | |
Db 2122 CTTTCACAG 2181
| | | | |
QY 1621 TGGGATATTTGGGCCCCAGAGCTTCAAGCTTCCCTTCAAGAGAGCTGATTTTGGC 1680
| | | | |
Db 2182 TGGGATATTTGGGCCCCAGAGCTTCAAGCTTCCCTTCAAGAGAGCTGATTTTGGC 2241
| | | | |
QY 1681 ACAGAGTCTCACTTACTCTGCTGAGCATCTAAGAGAGAGAGAGAGAGAGAGAGAGAG 1740
| | | | |

Db 2242 ACAGAGTCTCACTTACTTCTGCTCAGCCATCTACGAGGAGGAGCCAGTTACTT 2301
Qy 1741 GCCTACAGCTGACGCGAGCTGCCACTTTCGGGAGACCAAGTCTATTCTGTGCGAGGCG 1800
Db 2302 GCCTACAGCTGACGCGAGCTGCCACTTTCGGGAGACCAAGTCTATTCTGTGCGAGGCG 2361
Qy 1801 CAGAAAGCAAGTACAGAGCTGACTGCGCGGAGCTGGCATGGAAGAGGCCCCCTTGA 1860
Db 2362 CAGAAAGCAAGTACAGAGCTGACTGCGCGGAGCTGGCATGGAAGAGGCCCCCTTGA 2421
Qy 1861 AAGCGCTTAAACGAGCAAGCTGCCAATTTGAATTTGAGAGAGCATCATGTCAGAGAC 1920
Db 2422 AAGCGCTTAAACGAGCAAGCTGCCAATTTGAATTTGAGAGAGCATCATGTCAGAGAC 2481
Qy 1921 AGTCACGCGGAGAGCTGCGGAGAGTGGGCACTGCTACTTTTGGGCGAGCATGGA 1980
Db 2482 AGTCACGCGGAGAGCTGCGGAGAGTGGGCACTGCTACTTTTGGGCGAGCATGGA 2541
Qy 1981 ATCATTGAGGTCTCC 1995
Db 2542 ATCATTGAGGTCTCC 2556

RESULT 5
US-09-816-494-1
; Sequence 1, Application US/09816494
; Patent No. US20020034807A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel A.
; TITLE OF INVENTION: 38692 AND 21117, NOVEL DUAL SPECIFICITY
; TITLE OF INVENTION: PHOSPHATASE MOLECULES AND USES THEREFOR
; FILE REFERENCE: 10448-030002
; CURRENT APPLICATION NUMBER: US/09/816,494
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 60/191,858
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3544
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (589) ... (2583)
US-09-816-494-1

Query Match 94.9%; Score 1893; DB 9; Length 3544;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ATGGCCCATGAGTATGTTGAACTCAATTTGTTACTGAGAGGTGTGCTCTGCTGAA 60
Db 589 ATGGCCCATGAGTATGTTGAACTCAATTTGTTACTGAGAGGTGTGCTCTGCTGAA 648
Qy 61 AGTGAACGGAAGAAAGTGTCTGAATTTGATAGCGGCTTTTGTGAATTAATATCATCC 120
Db 649 AGTGAACGGAAGAAAGTGTCTGAATTTGATAGCGGCTTTTGTGAATTAATATCATCC 708
Qy 121 CACATTTTGAAGCCATTATATCACTGCTCAAGTTATATGAGCGAAGGTGCAAG 180
Db 709 CACATTTTGAAGCCATTATATCACTGCTCAAGTTATATGAGCGAAGGTGCAAG 768
Qy 181 GACAAAGTGTATATACAGAGCTCATCCAGATTAGCGAAACATTAAGTTGACATTGAT 240
Db 769 GACAAAGTGTATATACAGAGCTCATCCAGATTAGCGAAACATTAAGTTGACATTGAT 828
Qy 241 TGCAGTACGAAGGTGTAGTTAGCATCAAGCTCCGAAGGTGCTCTCTCTTTCA 300
Db 829 TGCAGTACGAAGGTGTAGTTAGCATCAAGCTCCGAAGGTGCTCTCTCTTTCA 888
Qy 301 GACTGTTTTCTCACTGTACTTCTGGGTAACTGAGAGAGCTTCAACTCTGTTCACTG 360

Db 889 GACTGTTTTCTCACTGTACTTCTGGGTAACTGAGAGAGCTTCAACTCTGTTCACTG 948
Qy 361 CTTCAGAGTGGGTTTTCAGTCTCTGTTGTTTCCCTGCGCTCTGTGAGAGAAATTC 420
Db 949 CTTCAGAGTGGGTTTTCAGTCTCTGTTGTTTCCCTGCGCTCTGTGAGAGAAATTC 1008
Qy 421 ACTTATGCTTCACTGATTTCTGAGCTTGTCTTACCTGTGTTGTTGTTGTTGTTGTT 480
Db 1009 ACTTATGCTTCACTGATTTCTGAGCTTGTCTTACCTGTGTTGTTGTTGTTGTTGTT 1068
Qy 481 CGAATTTCTCCAAATCTTATTTTGTGCTGCGAGAGATGTTCTCAACAGAGCTGATA 540
Db 1069 CGAATTTCTCCAAATCTTATTTTGTGCTGCGAGAGATGTTCTCAACAGAGCTGATA 1128
Qy 541 CAGCAAAATGAGTGTGTTATGTTAAATGCGAGTATACCTGTCCAAAGCTGACTT 600
Db 1129 CAGCAAAATGAGTGTGTTATGTTAAATGCGAGTATACCTGTCCAAAGCTGACTT 1188
Qy 601 ATCCCGAGTCTCAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 660
Db 1189 ATCCCGAGTCTCAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1248
Qy 661 CCGTGTGGAACAATCAGTATGTTTCAATGAGAAAGCAAAAGCTTCAATGATGTT 720
Db 1249 CCGTGTGGAACAATCAGTATGTTTCAATGAGAAAGCAAAAGCTTCAATGATGTT 1308
Qy 721 CTAGTCACTGTTTATGCTGAGTCTCCGCTCCGCAATGCTATGCTATGCTATGCTATG 780
Db 1309 CTAGTCACTGTTTATGCTGAGTCTCCGCTCCGCAATGCTATGCTATGCTATGCTATG 1368
Qy 781 AAGAGAGAGCAATGCTTTTATGATGAAGCTTACAGATTTTGAAGAAAAAGACTACT 840
Db 1369 AAGAGAGAGCAATGCTTTTATGATGAAGCTTACAGATTTTGAAGAAAAAGACTACT 1428
Qy 841 ATATCTCCAAATCTCAATTTTCTGCGCAACTCTGCTATGATGAGAAAGATTTAAGAAC 900
Db 1429 ATATCTCCAAATCTCAATTTTCTGCGCAACTCTGCTATGATGAGAAAGATTTAAGAAC 1488
Qy 901 CAGACTGAGATCAGGCGCAAGAGCAACTCAAGTGTGCTGCACTGAGAGCAACT 960
Db 1489 CAGACTGAGATCAGGCGCAAGAGCAACTCAAGTGTGCTGCACTGAGAGCAACT 1548
Qy 961 GAACCTGTCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
Db 1549 GAACCTGTCCCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1608
Qy 1021 TGTGCGACTCTGCTACTCCTGAGGAGGAGCAAGAGCCGCTGATCCGCGCAGCGTG 1080
Db 1609 TGTGCGACTCTGCTACTCCTGAGGAGGAGCAAGAGCCGCTGATCCGCGCAGCGTG 1668
Qy 1081 CCCAGGCTGCCAGCGCTGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
Db 1669 CCCAGGCTGCCAGCGCTGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1728
Qy 1141 AGTGGCTGCACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1200
Db 1729 AGTGGCTGCACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1788
Qy 1201 TCTCTGATATCAATCAAGTTTCAATTTCAATTCAGCAGCATGCGATCTTCAATGCTTC 1260
Db 1789 TCTCTGATATCAATCAAGTTTCAATTTCAATTCAGCAGCATGCGATCTTCAATGCTTC 1848
Qy 1261 TCTCTATCAGAAAGTGTGGAATATCACTTCACTATCTGTGATGGAACCAAC 1320
Db 1849 TCTCTATCAGAAAGTGTGGAATATCACTTCACTATCTGTGATGGAACCAAC 1908
Qy 1321 AAGCTATGCAAGTCTCCCTGTTTCAAGAACTATGAGAGAGAGTCCCGAAGCAAGTCT 1380
Db 1909 AAGCTATGCAAGTCTCCCTGTTTCAAGAACTATGAGAGAGAGTCCCGAAGCAAGTCT 1968
Qy 1381 GATTAAG 1440
Db 1969 GATTAAG 2028

Oy	1441	AGCAAGCCATTTCATTCGGCTCGAAGACACAGACAGAGGAGCCGCCAGAGGTCCTTTTA	1500
Db	2029	AGCAAGCCATTTCATTCGGCTCGAAGACACAGACAGAGGAGCCGCCAGAGGTCCTTTTA	2088
Oy	1501	TCTCCACTGCATCGAAGTGGAGCGTGGAGAGCAATTACACAGCCAGCTTCTTTTCGGC	1560
Db	2089	TCTCCACTGCATCGAAGTGGAGCGTGGAGAGCAATTACACAGCCAGCTTCTTTTCGGC	2148
Oy	1561	CTTTTCACACGCGACAGACAGCCTTCAGAACTCTGCTGGCTTGAAGGCTGGCAC	1620
Db	2149	CTTTTCACACGCGACAGACAGCCTTCAGAACTCTGCTGGCTTGAAGGCTGGCAC	2208
Oy	1621	TCGGATATCTTGGGCCCCCAGACCTCTACCCCTCCCTGACACGAGCAGCTGTATTGTGC	1680
Db	2209	TCGGATATCTTGGGCCCCCAGACCTCTACCCCTCCCTGACACGAGCAGCTGTATTGTGC	2268
Oy	1681	ACAGAGTCTCTACACTTCTACTCTGTCTGCTAGCCATCTACGAGGAGAGTCCAGTTACTCT	1740
Db	2269	ACAGAGTCTCTACACTTCTACTCTGTCTGCTAGCCATCTACGAGGAGAGTCCAGTTACTCT	2328
Oy	1741	GCCTACAGCTGCAAGCCAGCTGCCCACTTCCGGAGACCAAGTCTATTCTGTGGCAGGCGG	1800
Db	2329	GCCTACAGCTGCAAGCCAGCTGCCCACTTCCGGAGACCAAGTCTATTCTGTGGCAGGCGG	2388
Oy	1801	CAGAGGCCAAGTACAGACAGCTGACTCGCGGCGGAGCTGGCATGAGAGAGGCCCTTTGAA	1860
Db	2389	CAGAGGCCAAGTACAGACAGCTGACTCGCGGCGGAGCTGGCATGAGAGAGGCCCTTTGAA	2448
Oy	1861	AAGCAGTTTAAACGCGAAGAGCTGCCAATGGAATTTGGAGAGCATCATGTTCAGAAAC	1920
Db	2449	AAGCAGTTTAAACGCGAAGAGCTGCCAATGGAATTTGGAGAGCATCATGTTCAGAAAC	2508
Oy	1921	AGGTACCGGGAAGAGCTGGGGGAAAGTGGGCACTCAGTCTTCTTTCGGGACAGCATGGA	1980
Db	2509	AGGTACCGGGAAGAGCTGGGGGAAAGTGGGCACTCAGTCTTCTTTCGGGACAGCATGGA	2568
Oy	1981	ATCATTTAGAGTCTCC 1995	
Db	2569	ATCATTTAGAGTCTCC 2583	
RESULT 6			
US-10-377-072-25			
Sequence 25, Application US/10377072			
Publication No. US20040009501A1			
GENERAL INFORMATION:			
APPLICANT: Millennium Pharmaceuticals Inc.			
APPLICANT: Curtis, Rory A.J.			
APPLICANT: Logan, Thomas Joseph			
APPLICANT: Glucksmann, Maria A.			
APPLICANT: Meyers, Rachel E. J.			
APPLICANT: Williamson, Mark J.			
APPLICANT: Rudolph-Owen, Laura A.			
APPLICANT: Tsai, Feng-Ying			
TITLE OF INVENTION: NOVEL 25869, 25934, 26335, 50365, 21117,			
TITLE OF INVENTION: 38692, 46508, 16816, 16839, 49937, 49931 AND 49933 MOLECULES			
TITLE OF INVENTION: AND USES THEREFOR			
FILE REFERENCE: MP103-0180NM1M			
CURRENT FILING DATE: 2003-02-27			
PRIOR APPLICATION NUMBER: US 09/895,860			
PRIOR FILING DATE: 2001-06-29			
PRIOR APPLICATION NUMBER: US 60/215,370			
PRIOR FILING DATE: 2000-06-29			
PRIOR APPLICATION NUMBER: US 09/723,806			
PRIOR FILING DATE: 2000-11-28			
PRIOR APPLICATION NUMBER: US 60/187,455			
PRIOR FILING DATE: 2000-03-07			
PRIOR APPLICATION NUMBER: US 09/843,297			
PRIOR FILING DATE: 2001-04-25			
PRIOR APPLICATION NUMBER: US 60/199,801			

```

? PRIOR FILING DATE: 2000-04-26
? PRIOR APPLICATION NUMBER: US 09/861,801
? PRIOR FILING DATE: 2001-05-21
? PRIOR APPLICATION NUMBER: US 60/205,508
? PRIOR FILING DATE: 2000-05-19
? PRIOR APPLICATION NUMBER: US 09/816,494
? PRIOR FILING DATE: 2001-03-23
? PRIOR APPLICATION NUMBER: US 09/815,419
? PRIOR FILING DATE: 2001-03-22
? Remaining Prior Application data removed - See File Wrapper or PALM.
? NUMBER OF SEQ ID NOS: 114
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 25
? LENGTH: 3544
? TYPE: DNA
? ORGANISM: Homo Sapiens
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (589)...(2586)
US-10-377-072-25

```

Query Match	94.9%	Score 1893;	DB 16;	Length 3544;
Best Local Similarity	99.9%	Pred. No. 0;		
Matches 1993; Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0;

Oy	ATGGCCCATGAGATGTGAACTCAATATTTACTGAGAGTGTGGCTCTCTGGAA	60
Db	ATGGCCCATGAGATGTGAACTCAATATTTACTGAGAGTGTGGCTCTCTGGAA	589
Oy	AGTGAACGAAAAAGTCTGCTAATTGATAGCGGCCATTGTGGAAATACATACATCC	120
Db	AGTGAACGAAAAAGTCTGCTAATTGATAGCGGCCCATTTGTGAAATACATACATCC	649
Oy	CACATTTTGGAAAGCATTAATATCACTGCTCCAAAGCTATGAAAGCGAAGTTGCACAG	180
Db	CACATTTTGGAAAGCATTAATATCACTGCTCCAAAGCTATGAAAGCGAAGTTGCACAG	709
Oy	GACAAAGTGTAAATTAACAGAGCTCATCCAGCATTCAGCGAAACATAAGSTTGACATTGAT	240
Db	GACAAAGTGTAAATTAACAGAGCTCATCCAGCATTCAGCGAAACATAAGSTTGACATTGAT	769
Oy	TGCAAGTCAGAAAGTTGTAGTTACGATCAAAAGCTCCCAAGATGTGGCTCTCTCTTCA	300
Db	TGCAAGTCAGAAAGTTGTAGTTACGATCAAAAGCTCCCAAGATGTGGCTCTCTCTTCA	829
Oy	GACGTGTTTTCACCTGACTCTCTGGGTAACTGAGAAAGAGCTTCAACTGTTCACCTG	360
Db	GACGTGTTTTCACCTGACTCTCTGGGTAACTGAGAAAGAGCTTCAACTGTTCACCTG	889
Oy	CTTGACAGTGGGTTTGCTGAGTTCTCTGTTGTTTCCCTGGCCTCTGTGAAAGAAATCC	420
Db	CTTGACAGTGGGTTTGCTGAGTTCTCTGTTGTTTCCCTGGCCTCTGTGAAAGAAATCC	949
Oy	ACTCTAGTCCCTACCTGCAATTTCTCAGCCTTGCTTACCTGTGGCAATATGGGCCAACC	480
Db	ACTCTAGTCCCTACCTGCAATTTCTCAGCCTTGCTTACCTGTGGCAATATGGGCCAACC	1009
Oy	CGAATTCCTCCCAATCTTTATCTTGGCTGCAGCGAGATGTCTTCAACAGAGACTGATA	540
Db	CGAATTCCTCCCAATCTTTATCTTGGCTGCAGCGAGATGTCTTCAACAGAGACTGATA	1069
Oy	CAGCGAAATGGGATTTGGTTATGTGTAAATGCCCAGACTATACCTGTCCAAAGCTGACTTT	600
Db	CAGCGAAATGGGATTTGGTTATGTGTAAATGCCCAGACTATACCTGTCCAAAGCTGACTTT	1129
Oy	ATCCCGCAGTCTCATTTCTGCGCTGCTGCTGTGATGACAGCTTTTGTGAGAAATTTTG	660
Db	ATCCCGCAGTCTCATTTCTGCGCTGCTGCTGTGATGACAGCTTTTGTGAGAAATTTTG	1189
Oy	CCGTGTTTGGCAAAATCAGTAGATTTCAATTGAGAAAGCAAAAGCTCCAAATGATGTGTT	720
Db	CCGTGTTTGGCAAAATCAGTAGATTTCAATTGAGAAAGCAAAAGCTCCAAATGATGTGTT	1249

QY 721 CTAGTCACGTTTAACTGGAGATCTCCGCTCCGCCACACGCTATGCTTACATCATG 780
Db 1309 CTAGTSCACTGTTTAACTGGAGATCTCCGCTCCGCCACATGCTATGCTTACATCATG 1368
QY 781 AAGAGATGAGACATGCTTTAAGATGAAGCTTACAGATTTGTGAAGAAAAAGACTCT 840
Db 1369 AAGAGATGAGACATGCTTTAAGATGAAGCTTACAGATTTGTGAAGAAAAAGACTCT 1428
QY 841 ATATCTCCAACTTCAATTTTCTGGGCCAATCCCTGCACTATGAGAGAAATTAAGAAC 900
Db 1429 ATATCTCCAACTTCAATTTTCTGGGCCAATCCCTGCACTATGAGAGAAATTAAGAAC 1488
QY 901 CAGACTGAGACATCAGGGCCAAAGACAACTCAAGCTGTGCACTGTGAGAACCAAT 960
Db 1489 CAGACTGAGACATCAGGGCCAAAGACAACTCAAGCTGTGCACTGTGAGAACCAAT 1548
QY 961 GAACCTGTCCTGCTGCTTCAAGAGGTGACACAAAAAGCAGACGCCCTTCAAGTCCACC 1020
Db 1549 GAACCTGTCCTGCTGCTTCAAGAGGTGACACAAAAAGCAGACGCCCTTCAAGTCCACC 1608
QY 1021 TGTGCGCACTGTGCTACTCAGAGGACAGACGAAAGGCCCTGTCATCCCGCAGCTG 1080
Db 1609 TGTGCGCACTGTGCTACTCAGAGGACAGACGAAAGGCCCTGTCATCCCGCAGCTG 1668
QY 1081 CCCAGCGTCCCAAGCGCTGCAAGCCGTCCGTGTTAGAGGACAGCCCGCTGTATCAGAGCGCT 1140
Db 1669 CCCAGCGTCCCAAGCGCTGCAAGCCGTCCGTGTTAGAGGACAGCCCGCTGTATCAGAGCGCT 1728
QY 1141 AGTGGGCTGCACTGTGCTCCGACAGAGGCTGGAAGACAGCAATAAGCTCAAGCTTCTTC 1200
Db 1729 AGTGGGCTGCACTGTGCTCCGACAGAGGCTGGAAGACAGCAATAAGCTCAAGCTTCTTC 1788
QY 1201 TCTTGTGATATCAATATGTTTCATATTCAGCCAGCATGCGACATCTTACATGCTTC 1260
Db 1789 TCTTGTGATATCAATATGTTTCATATTCAGCCAGCATGCGACATCTTACATGCTTC 1848
QY 1261 TCCCTCATCAGAAAGTGTGTTGAATTAATAAATCTTCAATCTTGTGATGGAACCAAC 1320
Db 1849 TCCCTCATCAGAAAGTGTGTTGAATTAATAAATCTTCAATCTTGTGATGGAACCAAC 1908
QY 1321 AAGCTATGCCAGTTCCTCCCTGTTCAAGAACTATCGGAGAGACATCCGAAACCAAGTCT 1380
Db 1909 AAGCTATGCCAGTTCCTCCCTGTTCAAGAACTATCGGAGAGACATCCGAAACCAAGTCT 1968
QY 1381 GATTAAGAGAGAAACCCAGCATCCCAAGAAAGCTGCAAGCCGCAAGGCTTACAGACAGCAG 1440
Db 1969 GATTAAGAGAGAAACCCAGCATCCCAAGAAAGCTGCAAGCCGCAAGGCTTACAGACAGCAG 2028
QY 1441 AGCAAGCAGATTGCAATCGGTCAAGAACACAGAGAGAGGACCGCCCAAGAGTCCCTTTTA 1500
Db 2029 AGCAAGCAGATTGCAATCGGTCAAGAACACAGAGAGAGGACCGCCCAAGAGTCCCTTTTA 2088
QY 1501 TCTCCACTGATCGAAAGTGGAGAGCTGAGAGACAAATTAACAACCAAGCTTCTTTTGGGC 1560
Db 2089 TCTCCACTGATCGAAAGTGGAGAGCTGAGAGACAAATTAACAACCAAGCTTCTTTTGGGC 2148
QY 1561 CTTTCCACAGCAGAGAGACCTCAAGAAATGCTGTGGCTGTGGCTTTAAGGGCTGTGCAC 1620
Db 2149 CTTTCCACAGCAGAGAGACCTCAAGAAATGCTGTGGCTGTGGCTTTAAGGGCTGTGCAC 2208
QY 1621 TGGGATATTTGGCCCCCAAGACTCTTACCCCTTCCCTGACAGAGAGTGTATTTTGGC 1680
Db 2209 TGGGATATTTGGCCCCCAAGACTCTTACCCCTTCCCTGACAGAGAGTGTATTTTGGC 2268
QY 1681 ACAGAGTCTCAGACTTCTTACTCTGCTGACGACATCTTACGAGGACAGTCCAGTTACTCT 1740
Db 2269 ACAGAGTCTCAGACTTCTTACTCTGCTGACGACATCTTACGAGGACAGTCCAGTTACTCT 2328
QY 1741 GCTTACAGCTGAGCAGCTGCCACTTGGAGAGCAAGCTTATTTTGTGCGGAGGCGG 1800
Db 2329 GCTTACAGCTGAGCAGCTGCCACTTGGAGAGCAAGCTTATTTTGTGCGGAGGCGG 2388
QY 1801 CAGAAAGCAAGTGAAGCTGACTCGGCGGAGAGCTGGACATGAAGAGACCCCTTTGAA 1860

Db 2389 CAGAAAGCAAGTGAAGCTGACTCGGCGGAGAGCTGGACATGAAGAGACCCCTTTGAA 2448
QY 1861 AAGCAATTTAAACGCAAGAGCTGCCAATATGAAATTTGAGAGAGCATCATGTCAAGAAC 1920
Db 2449 AAGCAATTTAAACGCAAGAGCTGCCAATATGAAATTTGAGAGAGCATCATGTCAAGAAC 2508
QY 1921 AGTCCAGGGAAGAGCTGGGGAAGGTGGGACAGTCACTAGCTTTTGGGCGACATGGAA 1980
Db 2509 AGTCCAGGGAAGAGCTGGGGAAGGTGGGACAGTCACTAGCTTTTGGGCGACATGGAA 2568
QY 1981 ATCATTTAGAGTCTCC 1995
Db 2569 ATCATTTAGAGTCTCC 2583

RESULT 7
US-10-425-114-26234
; Sequence 26234, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kowalic, David K.
; APPLICANT: Screen, Steven B.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; NUMBER OF SEQ ID NOS: 2003-04-28
; SEQ ID NO 26234
; LENGTH: 3625
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB4119-028-HE_FLI
US-10-425-114-26234

Query Match 94.9%; Score 1893; DB 13; Length 3625;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGGCCCATGAGATGATTTGGAATCAATTTTACTGAGAGGTTGGTGTGCTGCTGAA 60
Db 692 ATGGCCCATGAGATGATTTGGAATCAATTTTACTGAGAGGTTGGTGTGCTGCTGAA 751
QY 61 AGTGAACGGAAAAGTGTGCTAATTTGATAGCCGCCATTTGTGGAATCAATACATCC 120
Db 752 AGTGAACGGAAAAGTGTGCTAATTTGATAGCCGCCATTTGTGGAATCAATACATCC 811
QY 121 CACATTTGGAAGCCATTAATATCACTGCTCAAGCTTATGAACGGAAGTTGCAACG 180
Db 812 CACATTTGGAAGCCATTAATATCACTGCTCAAGCTTATGAACGGAAGTTGCAACG 871
QY 181 GACAAAGGTAAATTAACAGAGCTATCAAGCATTTAGGAAACATTAAGTTGACATTGAT 240
Db 872 GACAAAGGTAAATTAACAGAGCTATCAAGCATTTAGGAAACATTAAGTTGACATTGAT 931
QY 241 TGCAGTCAGAAAGTGTGATTTTACATCAAAAGCTCCAAAGATGTTGCTCTCTCTTCA 300
Db 932 TGCAGTCAGAAAGTGTGATTTTACATCAAAAGCTCCAAAGATGTTGCTCTCTCTTCA 991
QY 301 GACTGTTTCTCACTGTACTTCTGGTAACTGGAAGAAAGACTTCAACTGTTCACCTG 360
Db 992 GACTGTTTCTCACTGTACTTCTGGTAACTGGAAGAAAGACTTCAACTGTTCACCTG 1051
QY 361 CTTGAGAGTGGATTTGCTGAGTCTCTGTTGTTTCCCTGGCTCTGTGAAAGAAATCC 420
Db 1052 CTTGAGAGTGGATTTGCTGAGTCTCTGTTGTTTCCCTGGCTCTGTGAAAGAAATCC 1111

QY 421 ACTGATGCTCCATGATCTTCTCAGCCCTTGCTTACCTGTGCAACATTGGGCCAAC 480
| | | | |
Db 1112 ACTGTGCTCCCTACCTGATCTTCTCAGCCCTTGCTTACCTGTGCAACATTGGGCCAAC 1171
| | | | |
QY 481 CGAATCTTCCCAATCTTATCTTGGCTGCGCAGAGAGTCTCTCAACAGAGCTGATA 540
| | | | |
Db 1172 CGAATCTTCCCAATCTTATCTTGGCTGCGCAGAGAGTCTCTCAACAGAGCTGATA 1231
| | | | |
QY 541 CAGCAGATGGAGTGGTATATGTGTTAATGCGAGCTATACCTGTGCCAAAGCTGACTT 600
| | | | |
Db 1232 CAGCAGATGGAGTGGTATATGTGTTAATGCGAGCAATACCTGTGCCAAAGCTGACTT 1291
| | | | |
QY 601 ATCCCGAGTCTAATTTCCGCGTGGCTGTGAAAGACGCTTTGTGGAATTTTG 660
| | | | |
Db 1292 ATCCCGAGTCTAATTTCCGCGTGGCTGTGAAAGACGCTTTGTGGAATTTTG 1351
| | | | |
QY 661 CCGTGGTTGGCAAAATCAGTAGATTTTCAATTGAGAAAGCAAAAGCTCTCAATGATGTT 720
| | | | |
Db 1352 CCGTGGTTGGCAAAATCAGTAGATTTTCAATTGAGAAAGCAAAAGCTCTCAATGATGTT 1411
| | | | |
QY 721 CTAGTGCACTGTTAGCTGGGATCTCCGCTCCGCCACATGCTATGCTTACATCATG 780
| | | | |
Db 1412 CTAGTGCACTGTTAGCTGGGATCTCCGCTCCGCCACATGCTTACATCATG 1471
| | | | |
QY 781 AAGAGATGGACATGTCCTTAGATGAAGCTTACAGATTGTGAAAGAAAGAAAGCTTACT 840
| | | | |
Db 1472 AAGAGATGGACATGTCCTTAGATGAAGCTTACAGATTGTGAAAGAAAGAAAGCTTACT 1531
| | | | |
QY 841 ATATCTCCAAACTTCAATTTTCTGGGCCAACTCTGTGACTATGAGAAAGATTAAAGAC 900
| | | | |
Db 1532 ATATCTCCAAACTTCAATTTTCTGGGCCAACTCTGTGACTATGAGAAAGATTAAAGAC 1591
| | | | |
QY 901 CAGACTGGAGCATCAGGCGCAAGAGCAAACTCAAGCTGTGCACTGTGAGAGCCAAAT 960
| | | | |
Db 1592 CAGACTGGAGCATCAGGCGCAAGAGCAAACTCAAGCTGTGCACTGTGAGAGCCAAAT 1651
| | | | |
QY 961 GAACTGTCCCTGCTCTCAGAGGCTGAGCAGAAAGAGAGCCCTCATCTCCACCC 1020
| | | | |
Db 1652 GAACTGTCCCTGCTCTCAGAGGCTGAGCAGAAAGAGAGCCCTCATCTCCACCC 1711
| | | | |
QY 1021 TGTGCGCACTCTGCTAAGAGGAGCAGAGCAAAAGGCGCTGATCCCGCAGCGTG 1080
| | | | |
Db 1712 TGTGCGCACTCTGCTAAGAGGAGCAGAGCAAAAGGCGCTGATCCCGCAGCGTG 1771
| | | | |
QY 1081 CCCAGGCTGCCAGCGCTGACCGCTGCTGTAGAGAGACGCCGCTGTGATCAGGCGCT 1140
| | | | |
Db 1772 CCCAGGCTGCCAGCGCTGACCGCTGCTGTAGAGAGACGCCGCTGTGATCAGGCGCT 1831
| | | | |
QY 1141 AGTGGCTGCACTGTCCGAGACAGGCTGGAAGACAGCAATTAAGCTCAAGCTTCTTC 1200
| | | | |
Db 1832 AGTGGCTGCACTGTCCGAGACAGGCTGGAAGACAGCAATTAAGCTCAAGCTTCTTC 1891
| | | | |
QY 1201 TCTCTGGATTCAAATCAGTTTATATTCAGCAGCAGTGGCAGCATCTTACATGCTTC 1260
| | | | |
Db 1892 TCTCTGGATTCAAATCAGTTTATATTCAGCAGCAGTGGCAGCATCTTACATGCTTC 1951
| | | | |
QY 1261 TCTCTCATCAGAGATGCTTGAATTAATACTAACCTTCACTCTGTGATGGAGCAAC 1320
| | | | |
Db 1952 TCTCTCATCAGAGATGCTTGAATTAATACTAACCTTCACTCTGTGATGGAGCAAC 2011
| | | | |
QY 1321 AAGCTATGCCAGTTTCTCCCTGTTCAGAGAACTATGCGAGAGACTCCGAAACCACTGCT 1380
| | | | |
Db 2012 AAGCTATGCCAGTTTCTCCCTGTTCAGAGAACTATGCGAGAGACTCCGAAACCACTGCT 2071
| | | | |
QY 1381 GATTAAGAGAGAGCCAGCATCCCAAGAGCTCAGACCGCGGCTTCAAGCAGCAG 1440
| | | | |
Db 2072 GATTAAGAGAGAGCCAGCATCCCAAGAGCTCAGACCGCGGCTTCAAGCAGCAG 2131
| | | | |
QY 1441 AGCAAGGATTTGATTCGATCAGAAACAGCAGAGTGGACCGGCCAGAGAGTCCCTTTTA 1500
| | | | |
Db 2132 AGCAAGGATTTGATTCGATCAGAAACAGCAGAGTGGACCGGCCAGAGAGTCCCTTTTA 2191
| | | | |
QY 1501 TCTCCACTGATGGAAGTGGAGCGTGGAGAGCAATTACACACGACTTCTTTTCGCG 1560
| | | | |

Db 2192 TCTCCACTGATGGAAGTGGAGCGTGGAGAGCAATTACACACGACTTCTTTTCGCG 2251
| | | | |
QY 1561 CTTTCCACAGCCAGCAGACAGCTTACAGAGTGTGCTGCGCTTGAAGGCTGGCAC 1620
| | | | |
Db 2252 CTTTCCACAGCCAGCAGACAGCTTACAGAGTGTGCTGCGCTTGAAGGCTGGCAC 2311
| | | | |
QY 1621 TCGGATATCTTGGCCCCCGACCTTACCCCTTCCCTTACACAGCAGCTGTATTTTGGC 1680
| | | | |
Db 2312 TCGGATATCTTGGCCCCCGACCTTACCCCTTCCCTTACACAGCAGCTGTATTTTGGC 2371
| | | | |
QY 1681 ACAGAGTCTTCAACTTCTACTCTGCGCTCAGCAGCATTAAGAGGAGCTGAGTACTCT 1740
| | | | |
Db 2372 ACAGAGTCTTCAACTTCTACTCTGCGCTCAGCAGCATTAAGAGGAGCTGAGTACTCT 2431
| | | | |
QY 1741 GCTTACAGCTGACAGCCAGCTGCCCATTTGCGAGACCAAGTATTTCTGTGCGCAGCGG 1800
| | | | |
Db 2432 GCTTACAGCTGACAGCCAGCTGCCCATTTGCGAGACCAAGTATTTCTGTGCGCAGCGG 2491
| | | | |
QY 1801 CAGAGCCAGTGGACAGAGCTGATCTGCGCGGAGCTGGCATGAAAGAGCCCTTTGAA 1860
| | | | |
Db 2492 CAGAGCCAGTGGACAGAGCTGATCTGCGCGGAGCTGGCATGAAAGAGCCCTTTGAA 2551
| | | | |
QY 1861 AAGCAGTTTAAACGAGAGCTGCCAAATGGAATTTGAGAGAGCATGTGAGAGAAC 1920
| | | | |
Db 2552 AAGCAGTTTAAACGAGAGCTGCCAAATGGAATTTGAGAGAGCATGTGAGAGAAC 2611
| | | | |
QY 1921 AGGTCCGGAAGAGCTGGGGAAGTGGGAGTCACTAGCTTTTGGGCGAGCATGAA 1980
| | | | |
Db 2612 AGGTCCGGAAGAGCTGGGGAAGTGGGAGTCACTAGCTTTTGGGCGAGCATGAA 2671
| | | | |
QY 1981 ATCATTGAGCTTCC 1995
| | | | |
Db 2672 ATCATTGAGCTTCC 2686
| | | | |

RESULT 8
US-10-343-357-17
; Sequence 17, Application US/10343357
; Publication No. US20040058341A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.; TANG, Y.Tom
; APPLICANT: ELIOTT, Vicki S.; RAMKUMAR, Jayalaxmi
; APPLICANT: YAO, Monique G.; BURFORD, Neil
; APPLICANT: WANG, Yumei E.; STEWART, Elizabeth A.
; APPLICANT: GANDHI, Ameena R.; ARVIZU, Chandra S.
; APPLICANT: LEE, Ernestine A.; HAPALIA, April J.A.
; APPLICANT: LU, Dying Aina M.; TRIBOULEY, Catherine M.
; APPLICANT: GRIFFIN, Jennifer A.; BAUGHN, Mariah R.
; APPLICANT: YUE, Henry; WARREN, Bridget A.
; APPLICANT: NGUYEN, Daniel B.; CHAWLA, Nandinder K.
; APPLICANT: KEARNEY, Liam
; TITLE OF INVENTION: PROTEIN PHOSPHATASES
; FILE REFERENCE: PI-0173 PCT
; CURRENT APPLICATION NUMBER: US/10/343,357
; CURRENT FILING DATE: 2003-01-28
; PRIOR APPLICATION NUMBER: PCT/US01/23716
; PRIOR FILING DATE: 2001-07-26
; PRIOR APPLICATION NUMBER: US 60/221,679
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: US 60/223,272
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/224,309
; PRIOR FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: US 60/226,728
; PRIOR FILING DATE: 2000-08-18
; PRIOR APPLICATION NUMBER: US 60/229,254
; PRIOR FILING DATE: 2000-08-30
; PRIOR APPLICATION NUMBER: US 60/231,366
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PERL Program
; SEQ ID NO 17

LENGTH: 3766
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20040058341A1 7480570CB1
US-10-343-357-17

Query Match 94.9%; Score 1893; DB 13; Length 3766;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGGCCCATGAGATGATGGAATCTCAATTGTTACTGAGAGGTGGTGGCTCTGCTGGA 60
Db ATGGCCCATGAGATGATGGAATCTCAATTGTTACTGAGAGGTGGTGGCTCTGCTGGA 597
QY 61 AGTGAACGGAAGAAAGTCTGCTAATTGATAGCCGGCCATTGTTGGAAATACATATACC 120
Db AGTGAACGGAAGAAAGTCTGCTAATTGATAGCCGGCCATTGTTGGAAATACATATACC 657
QY 121 CACATTTTGAAGCCATTAATATATCACTGCTCAAGCTTAATGAGCGAAGTTGCACAG 180
Db CACATTTTGAAGCCATTAATATATCACTGCTCAAGCTTAATGAGCGAAGTTGCACAG 717
QY 658 CACATTTTGAAGCCATTAATATATCACTGCTCAAGCTTAATGAGCGAAGTTGCACAG 180
Db CACATTTTGAAGCCATTAATATATCACTGCTCAAGCTTAATGAGCGAAGTTGCACAG 717
QY 181 GACAAAGCTTAATTAACAGAGCTCATCCAGCATTCAGCGAACAATTAAGTTGACATTGAT 240
Db GACAAAGCTTAATTAACAGAGCTCATCCAGCATTCAGCGAACAATTAAGTTGACATTGAT 777
QY 241 TGCAGTCAGAAAGTTGATGATTAAGATCAAAAGTCCCAAGATGTTGCTCTCTCTCA 300
Db TGCAGTCAGAAAGTTGATGATTAAGATCAAAAGTCCCAAGATGTTGCTCTCTCTCA 300
QY 778 TGCAGTCAGAAAGTTGATGATTAAGATCAAAAGTCCCAAGATGTTGCTCTCTCTCA 837
Db TGCAGTCAGAAAGTTGATGATTAAGATCAAAAGTCCCAAGATGTTGCTCTCTCTCA 837
QY 301 GACTGTTTCTCACTGATCTTCTGGATTAACCTGAGAAAGCTTCAACTCTGTTACCTG 360
Db GACTGTTTCTCACTGATCTTCTGGATTAACCTGAGAAAGCTTCAACTCTGTTACCTG 897
QY 361 CTTCAGAGTGGGTTTGGTGAAGTCTCTGCTGTTTCCCTGGCTCTGTGAAGAAATCC 420
Db CTTCAGAGTGGGTTTGGTGAAGTCTCTGCTGTTTCCCTGGCTCTGTGAAGAAATCC 957
QY 898 CTTCAGAGTGGGTTTGGTGAAGTCTCTGCTGTTTCCCTGGCTCTGTGAAGAAATCC 957
Db CTTCAGAGTGGGTTTGGTGAAGTCTCTGCTGTTTCCCTGGCTCTGTGAAGAAATCC 957
QY 421 ACTCTAGTCCCTAAGTCTGATCTTCTGAGCTTCTGCTGTTTCCCTGGCTCTGTGAAGAAATCC 480
Db ACTCTAGTCCCTAAGTCTGATCTTCTGAGCTTCTGCTGTTTCCCTGGCTCTGTGAAGAAATCC 957
QY 958 ACTCTAGTCCCTAAGTCTGATCTTCTGAGCTTCTGCTGTTTCCCTGGCTCTGTGAAGAAATCC 480
Db ACTCTAGTCCCTAAGTCTGATCTTCTGAGCTTCTGCTGTTTCCCTGGCTCTGTGAAGAAATCC 957
QY 481 CGAATTTCTCCATCTTATCTTGGCTGCGACGAGATGCTCTCAACAGAGCTGTATA 540
Db CGAATTTCTCCATCTTATCTTGGCTGCGACGAGATGCTCTCAACAGAGCTGTATA 1017
QY 1018 CGAATTTCTCCATCTTATCTTGGCTGCGACGAGATGCTCTCAACAGAGCTGTATA 540
Db CGAATTTCTCCATCTTATCTTGGCTGCGACGAGATGCTCTCAACAGAGCTGTATA 1077
QY 541 CAGCAAAATGGAGTTGATGTTAATGCAAGCTATACCTGTCCAAAGCTGATCTT 600
Db CAGCAAAATGGAGTTGATGTTAATGCAAGCTATACCTGTCCAAAGCTGATCTT 1137
QY 1078 CAGCAAAATGGAGTTGATGTTAATGCAAGCTATACCTGTCCAAAGCTGATCTT 600
Db CAGCAAAATGGAGTTGATGTTAATGCAAGCTATACCTGTCCAAAGCTGATCTT 1137
QY 601 ATCCCGAGTCTCATTTCTGCGTGTGCTGTGAATGACAGCTTTTGTGAAGAAATTTTG 660
Db ATCCCGAGTCTCATTTCTGCGTGTGCTGTGAATGACAGCTTTTGTGAAGAAATTTTG 1197
QY 1138 ATCCCGAGTCTCATTTCTGCGTGTGCTGTGAATGACAGCTTTTGTGAAGAAATTTTG 660
Db ATCCCGAGTCTCATTTCTGCGTGTGCTGTGAATGACAGCTTTTGTGAAGAAATTTTG 1197
QY 661 CCGTGGTTGGCAAAATCACTATGATTGAGAAAGCAAAAGCTTCCAAATGATGTT 720
Db CCGTGGTTGGCAAAATCACTATGATTGAGAAAGCAAAAGCTTCCAAATGATGTT 1257
QY 1198 CCGTGGTTGGCAAAATCACTATGATTGAGAAAGCAAAAGCTTCCAAATGATGTT 720
Db CCGTGGTTGGCAAAATCACTATGATTGAGAAAGCAAAAGCTTCCAAATGATGTT 1257
QY 721 CTAGTGAATCTGTTAGTGGAGTCTCCGCTCCGACACATGCGTATGCGCTTCAATCATG 780
Db CTAGTGAATCTGTTAGTGGAGTCTCCGCTCCGACACATGCGTATGCGCTTCAATCATG 1317
QY 1258 CTAGTGAATCTGTTAGTGGAGTCTCCGCTCCGACACATGCGTATGCGCTTCAATCATG 780
Db CTAGTGAATCTGTTAGTGGAGTCTCCGCTCCGACACATGCGTATGCGCTTCAATCATG 1317
QY 781 AAGAGATGAGCATGCTTTAGATGAAGCTTAAGATTTGTGAAGAAAGAAAGCTTACT 840
Db AAGAGATGAGCATGCTTTAGATGAAGCTTAAGATTTGTGAAGAAAGAAAGCTTACT 1377
QY 1318 AAGAGATGAGCATGCTTTAGATGAAGCTTAAGATTTGTGAAGAAAGAAAGCTTACT 840
Db AAGAGATGAGCATGCTTTAGATGAAGCTTAAGATTTGTGAAGAAAGAAAGCTTACT 1377
QY 841 ATATCTCCAAATCTCAATTTTCTGGGCAACTCTCTGAGCTATGAGAAAGATTTAAGAAC 900
Db ATATCTCCAAATCTCAATTTTCTGGGCAACTCTCTGAGCTATGAGAAAGATTTAAGAAC 1437
QY 1378 ATATCTCCAAATCTCAATTTTCTGGGCAACTCTCTGAGCTATGAGAAAGATTTAAGAAC 900
Db ATATCTCCAAATCTCAATTTTCTGGGCAACTCTCTGAGCTATGAGAAAGATTTAAGAAC 1437
QY 901 CAGACTGAGATGAGGCGCAAAAGCAAACTCAAGCTGTGACCTGAGAAAGCAAAAT 960
Db CAGACTGAGATGAGGCGCAAAAGCAAACTCAAGCTGTGACCTGAGAAAGCAAAAT 960

Db 1438 CAGACTGAGATGAGGCGCAAAAGCAAACTCAAGCTGTGACCTGAGAAAGCAAAAT 1497
QY 961 GAACCTGTCCCTGCTCTCTCAAGAGGTGACAGAAAGCGAGAGCCCTCAGTCCACCC 1020
Db GAACCTGTCCCTGCTCTCTCAAGAGGTGACAGAAAGCGAGAGCCCTCAGTCCACCC 1557
QY 1021 TGTGCGCATCTGCTAATCTCAAGAGCGAGAGCAAAAGCCCGTCAATCCCGCAGCGTG 1080
Db TGTGCGCATCTGCTAATCTCAAGAGCGAGAGCAAAAGCCCGTCAATCCCGCAGCGTG 1617
QY 1558 TGTGCGCATCTGCTAATCTCAAGAGCGAGAGCAAAAGCCCGTCAATCCCGCAGCGTG 1080
Db TGTGCGCATCTGCTAATCTCAAGAGCGAGAGCAAAAGCCCGTCAATCCCGCAGCGTG 1617
QY 1081 CCCAGGTGCCCCAGGTGAGCGCTGCTGTTAGAGCAAGCCCGTGTACAGCGCTC 1140
Db CCCAGGTGCCCCAGGTGAGCGCTGCTGTTAGAGCAAGCCCGTGTACAGCGCTC 1677
QY 1141 AGTGGGCTGCACTGCTCCGAGAGAGGTGAGAGCAGCAATTAAGCTTAAGCTTCCCTTC 1200
Db AGTGGGCTGCACTGCTCCGAGAGAGGTGAGAGCAGCAATTAAGCTTAAGCTTCCCTTC 1737
QY 1678 AGTGGGCTGCACTGCTCCGAGAGAGGTGAGAGCAGCAATTAAGCTTCCCTTC 1200
Db AGTGGGCTGCACTGCTCCGAGAGAGGTGAGAGCAGCAATTAAGCTTCCCTTC 1737
QY 1201 TCTGGAATTAACAATGATTTCATATTCAGCCAGATGAGAGCTTCAATGAGCTTC 1260
Db TCTGGAATTAACAATGATTTCATATTCAGCCAGATGAGAGCTTCAATGAGCTTC 1797
QY 1261 TCTGGAATTAACAATGATTTCATATTCAGCCAGATGAGAGCTTCAATGAGCTTC 1260
Db TCTGGAATTAACAATGATTTCATATTCAGCCAGATGAGAGCTTCAATGAGCTTC 1797
QY 1798 TCTGGAATTAACAATGATTTCATATTCAGCCAGATGAGAGCTTCAATGAGCTTC 1260
Db TCTGGAATTAACAATGATTTCATATTCAGCCAGATGAGAGCTTCAATGAGCTTC 1797
QY 1321 AAGCTATGCAATGCTCCCTGTTCAAGAACTATGAGAGAGTCCCGAAACAGTCT 1380
Db AAGCTATGCAATGCTCCCTGTTCAAGAACTATGAGAGAGTCCCGAAACAGTCT 1917
QY 1858 AAGCTATGCAATGCTCCCTGTTCAAGAACTATGAGAGAGTCCCGAAACAGTCT 1380
Db AAGCTATGCAATGCTCCCTGTTCAAGAACTATGAGAGAGTCCCGAAACAGTCT 1917
QY 1381 GATTAAGAGAGAGCGAGATGCTCCAGAAAGCTGAGAGCGGCTTCAAGAGCAG 1440
Db GATTAAGAGAGAGCGAGATGCTCCAGAAAGCTGAGAGCGGCTTCAAGAGCAG 1977
QY 1918 GATTAAGAGAGAGCGAGATGCTCCAGAAAGCTGAGAGCGGCTTCAAGAGCAG 1440
Db GATTAAGAGAGAGCGAGATGCTCCAGAAAGCTGAGAGCGGCTTCAAGAGCAG 1977
QY 1441 AGCAAGCGATGCTCCCTGTTCAAGAACTATGAGAGAGTCCCGAAACAGTCT 1500
Db AGCAAGCGATGCTCCCTGTTCAAGAACTATGAGAGAGTCCCGAAACAGTCT 2037
QY 1978 AGCAAGCGATGCTCCCTGTTCAAGAACTATGAGAGAGTCCCGAAACAGTCT 1500
Db AGCAAGCGATGCTCCCTGTTCAAGAACTATGAGAGAGTCCCGAAACAGTCT 2037
QY 1501 TCTCACTGATGCAAGTGGAGAGGTGAGAGCAATTAACCAACAGCTTCTTTCCGG 1560
Db TCTCACTGATGCAAGTGGAGAGGTGAGAGCAATTAACCAACAGCTTCTTTCCGG 2097
QY 1561 CTTCACACAGCGAGAGCTTCAAGAACTGCTGAGCTTCTTTCCGG 1620
Db CTTCACACAGCGAGAGCTTCAAGAACTGCTGAGCTTCTTTCCGG 2157
QY 2098 CTTCACACAGCGAGAGCTTCAAGAACTGCTGAGCTTCTTTCCGG 1620
Db CTTCACACAGCGAGAGCTTCAAGAACTGCTGAGCTTCTTTCCGG 2157
QY 1621 TCGGATATCTGGCCCCCAGACCTCTACCCCTTCCCTGACAGAGCTGTTATTTGCG 1680
Db TCGGATATCTGGCCCCCAGACCTCTACCCCTTCCCTGACAGAGCTGTTATTTGCG 2217
QY 2158 TCGGATATCTGGCCCCCAGACCTCTACCCCTTCCCTGACAGAGCTGTTATTTGCG 1680
Db TCGGATATCTGGCCCCCAGACCTCTACCCCTTCCCTGACAGAGCTGTTATTTGCG 2217
QY 1681 ACAGAGTCTCTCAACATCTTACTGCTGCTCAAGCCATGAGAGCGAGTCCAGTTACTCT 1740
Db ACAGAGTCTCTCAACATCTTACTGCTGCTCAAGCCATGAGAGCGAGTCCAGTTACTCT 2277
QY 2218 ACAGAGTCTCTCAACATCTTACTGCTGCTCAAGCCATGAGAGCGAGTCCAGTTACTCT 1740
Db ACAGAGTCTCTCAACATCTTACTGCTGCTCAAGCCATGAGAGCGAGTCCAGTTACTCT 2277
QY 1741 GCTTACAGCTGAGCGAGCGCCACTTGGAGAGCAAGTCTTATTTCTGTGGCGGCG 1800
Db GCTTACAGCTGAGCGAGCGCCACTTGGAGAGCAAGTCTTATTTCTGTGGCGGCG 2337
QY 2278 GCTTACAGCTGAGCGAGCGCCACTTGGAGAGCAAGTCTTATTTCTGTGGCGGCG 1800
Db GCTTACAGCTGAGCGAGCGCCACTTGGAGAGCAAGTCTTATTTCTGTGGCGGCG 2337
QY 1801 CAGAAAGCAAGTGAAGAGCTGACTGCGCGAGCTGAGAGAGAGAGAGAGAGAGAGAGAG 1860
Db CAGAAAGCAAGTGAAGAGCTGACTGCGCGAGCTGAGAGAGAGAGAGAGAGAGAGAGAG 2397
QY 2338 CAGAAAGCAAGTGAAGAGCTGACTGCGCGAGCTGAGAGAGAGAGAGAGAGAGAGAG 1860
Db CAGAAAGCAAGTGAAGAGCTGACTGCGCGAGCTGAGAGAGAGAGAGAGAGAGAGAGAG 2457
QY 1861 AAGCAAGTTTAAACGAGAGAGCTGCAATTAAGATTTGAGAGAGAGATCATGTCAAGAAC 1920
Db AAGCAAGTTTAAACGAGAGAGCTGCAATTAAGATTTGAGAGAGAGATCATGTCAAGAAC 2457
QY 2398 AAGCAAGTTTAAACGAGAGAGCTGCAATTAAGATTTGAGAGAGAGATCATGTCAAGAAC 1920
Db AAGCAAGTTTAAACGAGAGAGCTGCAATTAAGATTTGAGAGAGAGATCATGTCAAGAAC 2457
QY 1921 AAGTCAAGGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
Db AAGTCAAGGAGAGAGCTGAG 2517
QY 2458 AAGTCAAGGAGAGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
Db AAGTCAAGGAGAGAGCTGAG 2517
QY 1981 ATCATTTGAGGTCTCC 1995
Db ATCATTTGAGGTCTCC 1995

```
Db      2518 ATCATTGAGGTCTCC 2532

RESULT 9
US-10-648-593-115
; Sequence 115, Application US/10648593
; Publication No. US20040106132A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: IDENTIFICATION OF GENES FOR PREDICTING ACTIVITY OF COMPOUNDS THAT
; INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE KINASES AND/OR
; FILE REFERENCE: D0273 NP
; CURRENT APPLICATION NUMBER: US/10/648,593
; PRIOR FILING DATE: 2003-08-26
; PRIOR APPLICATION NUMBER: 60/406,385
; NUMBER OF SEQ ID NOS: 557
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 115
; LENGTH: 4790
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-648-593-115

Query Match      94.98; Score 1893; DB 17; Length 4790;
Best Local Similarity 99.98; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 ATGCCCCATGAGATGATTGGAAGCTCAAAATGTTACTGAGAGGTTGGTGGCTCTGCTGGAA 60
Db      184 ATGCCCCATGAGATGATTGGAAGCTCAAAATGTTACTGAGAGGTTGGTGGCTCTGCTGGAA 243
QY      61 AGTGAACGGAAGAGTGTCTAATGATAGCCGGCCATTGTGGAATACAAATACATCC 120
Db      244 AGTGAACGGAAGAGTGTCTAATGATAGCCGGCCATTGTGGAATACAAATACATCC 303
QY      121 CACATTTTGAAGCCATTATATCACTGCTCCAACTTATGAAGCAAGCTTGCAACAG 180
Db      304 CACATTTTGAAGCCATTATATCACTGCTCCAACTTATGAAGCAAGCTTGCAACAG 363
QY      181 GACAAAGTGTAAATTACAGAGCTCATCCAGCATTCAGCGAAGCAATTAAGTTGACATTGAT 240
Db      364 GACAAAGTGTAAATTACAGAGCTCATCCAGCATTCAGCGAAGCAATTAAGTTGACATTGAT 423
QY      241 TGAAGTCAGAAAGTTGATGTTTACGATCAAAAGCTCCCAAGATGTTGCTCTCTCTTCA 300
Db      424 TGAAGTCAGAAAGTTGATGTTTACGATCAAAAGCTCCCAAGATGTTGCTCTCTCTTCA 483
QY      301 GACTGTTTCTCACTGTAATCTTGCGTAACCTGGAAGAGCTTCAACTCTGTTCACTTG 360
Db      484 GACTGTTTCTCACTGTAATCTTGCGTAACCTGGAAGAGCTTCAACTCTGTTCACTTG 543
QY      361 CTGCAAGTGGGTTTGTGAGTCTCTCGTTGTTTCCCTGGCCCTGTGGAAGGAATATCC 420
Db      544 CTGCAAGTGGGTTTGTGAGTCTCTCGTTGTTTCCCTGGCCCTGTGGAAGGAATATCC 603
QY      421 ACTCTAGTCCCTACCTGATCTTCTCAGCCTTGCTTACCTGTTGCAAACTTTGGCCCAAC 480
Db      604 ACTCTAGTCCCTACCTGATCTTCTCAGCCTTGCTTACCTGTTGCAAACTTTGGCCCAAC 663
QY      481 CGAATTTCTCCCAATCTTTATCTTGGCTGCCAGAGAGATGCTCTCAAGAGAGCTGATA 540
Db      664 CGAATTTCTCCCAATCTTTATCTTGGCTGCCAGAGAGATGCTCTCAAGAGAGCTGATA 723
QY      541 CAGAGAGATGGGATGTTGTTATGTTTAATGCGACGATACCTGTCCAAAGCCGTGACTT 660
Db      724 CAGAGAGATGGGATGTTGTTATGTTTAATGCGACGATACCTGTCCAAAGCCGTGACTT 783
QY      601 ATCCCGAGTCTCATTTCTGCGGTGCTGCTGTAATGACAGCTTTTGTGAAAAATTTTG 660
Db      784 ATCCCGAGTCTCATTTCTGCGGTGCTGCTGTAATGACAGCTTTTGTGAAAAATTTTG 843

QY      661 CCGTGGTTGGAACAATCAGATGATTTCAATTGAGAAAGCAAAAGCTCCATGATGATGT 720
Db      844 CCGTGGTTGGAACAATCAGATGATTTCAATTGAGAAAGCAAAAGCTCCATGATGATGT 903
QY      721 CTAGTGCACTGTTTAACTGGGATCTCCCGCTCCGCCACCATGCTATTCGCTTACATATG 780
Db      904 CTAGTGCACTGTTTAACTGGGATCTCCCGCTCCGCCACCATGCTATTCGCTTACATATG 963
QY      781 AAGAGATGACATGCTTTTAAATGAGCTTACAGATTTTGTGAAGAAAAAAGAACTTACT 840
Db      964 AAGAGATGACATGCTTTTAAATGAGCTTACAGATTTTGTGAAGAAAAAAGAACTTACT 1023
QY      841 ATATCCCAACTTCAATTTTTCGGGCAACTCTGAGCTATGAGAAAGATTAAGAAC 900
Db      1024 ATATCCCAACTTCAATTTTTCGGGCAACTCTGAGCTATGAGAAAGATTAAGAAC 1083
QY      901 CAGACTGAGCATCAGAGGCCAAAGCAAACTCAAGCTGCTGACCTGAGAGAACCAAT 960
Db      1084 CAGACTGAGCATCAGAGGCCAAAGCAAACTCAAGCTGCTGACCTGAGAGAACCAAT 1143
QY      961 GAACTGTCTCTGCTGTCTCAGAGGTGACAGAAAAAGGAGAGCCCTCACTGATCCACCC 1020
Db      1144 GAACTGTCTCTGCTGTCTCAGAGGTGACAGAAAAAGGAGAGCCCTCACTGATCCACCC 1203
QY      1021 TGTGCGGACTTGTCTCTCAGAGGAGAGAGACAAAGCCGTGATCCCGCAGCGTG 1080
Db      1204 TGTGCGGACTTGTCTCTCAGAGGAGAGAGACAAAGCCGTGATCCCGCAGCGTG 1263
QY      1081 CCCAGCGTCCCGCAGCGTGTGAGCGGTGTTAGAGAGACAGCCGCTGTTAGAGGCGTCT 1140
Db      1264 CCCAGCGTCCCGCAGCGTGTGAGCGGTGTTAGAGAGACAGCCGCTGTTAGAGGCGTCT 1323
QY      1141 AGTGGCTGCACTGTCTCCGACAGAGCTGGAAGACAGCAATTAAGCTTCAAGCTTCTTC 1200
Db      1324 AGTGGCTGCACTGTCTCCGACAGAGCTGGAAGACAGCAATTAAGCTTCAAGCTTCTTC 1383
QY      1201 TCTCTGATATCAAAACAGTTTCAATTCATTCAGCCAGATGAGCAATCTTACATGCTTC 1260
Db      1384 TCTCTGATATCAAAACAGTTTCAATTCATTCAGCCAGATGAGCAATCTTACATGCTTC 1443
QY      1261 TCCTCATCAGAAAGATCTTTGGAATCTTCAAACTTCTCACTACTGTGATGGAACCAAC 1320
Db      1444 TCCTCATCAGAAAGATCTTTGGAATCTTCAAACTTCTCACTACTGTGATGGAACCAAC 1503
QY      1321 AAGCTATGCAAGTGTCTCCCTGTTTCAAGAACTATGAGACAGATCTCCGAAACCAAGTCT 1380
Db      1504 AAGCTATGCAAGTGTCTCCCTGTTTCAAGAACTATGAGACAGATCTCCGAAACCAAGTCT 1563
QY      1381 GATTAAGAGAAAGCCAGCATCTCCCAAGAAAGCTGACACCGCCAGGCTTCAACACCCAG 1440
Db      1564 GATTAAGAGAAAGCCAGCATCTCCCAAGAAAGCTGACACCGCCAGGCTTCAACACCCAG 1623
QY      1441 AGCAAGCAATTTGATTCGGTCAAGAACCAAGAGAGTGGCACGCGCCAGAGAGTCCCTTTTA 1500
Db      1624 AGCAAGCAATTTGATTCGGTCAAGAACCAAGAGAGTGGCACGCGCCAGAGAGTCCCTTTTA 1683
QY      1501 TCTTCACATCATGAAAGTGGAGACCTGAGAGCAATTAACACACAGCTTCTTTTCGCG 1560
Db      1684 TCTTCACATCATGAAAGTGGAGACCTGAGAGCAATTAACACACAGCTTCTTTTCGCG 1743
QY      1561 CTTTCCACAGCCAGAGACACTCAGAAAGTCTGCTGGCTGGGCTTTAAGGGCTGGCAC 1620
Db      1744 CTTTCCACAGCCAGAGACACTCAGAAAGTCTGCTGGCTGGGCTTTAAGGGCTGGCAC 1803
QY      1621 TCGGATATCTTGGCCCCCAGACCTCTACCCCTTCCCTGACACGACGCTGTATTTTGGC 1680
Db      1804 TCGGATATCTTGGCCCCCAGACCTCTACCCCTTCCCTGACACGACGCTGTATTTTGGC 1863
QY      1681 ACAGAGTCTCTACACTTCTACTGTGCTGAGCCATTAAGAGAGAGAGTCCAGTTACTCT 1740
Db      1864 ACAGAGTCTCTACACTTCTACTGTGCTGAGCCATTAAGAGAGAGAGTCCAGTTACTCT 1923
QY      1741 GCCTACAGCTGACAGCCAGCTGCCACTTGGCGAGACCAAGTCTATTTCTGTGCGCAGCGG 1800
```


Db 1924 GCGTACAGCTGCGAGCTGCGCACTTGGGAGACCAAGTCTATTCTGTGGCGAGCGG 1983
QY 1801 CAGAGCGAAGTGAAGAGAGTGACTGCGCGCGAGCTGGCATGAAGAGCCCTTTGAA 1860
Db 1984 CAGAGCGAAGTGAAGAGAGTGACTGCGCGCGAGCTGGCATGAAGAGCCCTTTGAA 2043
QY 1861 AAGCAGTTTAAACGCGAAGAGTGCCTAATGAAATTGGAGAGCATATGTCAGAGAAC 1920
Db 2044 AAGCAGTTTAAACGCGAAGAGTGCCTAATGAAATTGGAGAGCATATGTCAGAGAAC 2103
QY 1921 AGGTACCGGGAAGAGTGGGGAAGTGGGAGTCACTGCTTTCGGGAGAGTGA 1980
Db 2104 AGGTACCGGGAAGAGTGGGGAAGTGGGAGTCACTGCTTTCGGGAGAGTGA 2163
QY 1981 ATCATGAGAGTCTCC 1995
Db 2164 ATCATGAGAGTCTCC 2178

RESULT 10

US-10-296-115-520
; Sequence 520, Application US/10296115
; Publication No. US20040053248A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq Inc
; TITLE OF INVENTION: No. US20040053248A1el Nucleic Acids and Polypeptides
; FILE REFERENCE: 784PCT
; CURRENT FILING DATE: 2002-11-18
; PRIOR FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: US09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US09/552,317
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 1478
; SEQ ID NO 520
; LENGTH: 2966
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(2966)
; OTHER INFORMATION: n = a,t,c or g
US-10-296-115-520

Query Match 92.7%; Score 1849; DB 13; Length 2966;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1949; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 45 GGTGGCTCTGCTGGAAGTGAACGGAAGAGTCTGCTAATTTGATAGCCGCCATTGT 104
Db 67 GGTGGCTCTGCTGGAAGTGAACGGAAGAGTCTGCTAATTTGATAGCCGCCATTGT 126
QY 105 GGAATTCATTAATCCCAATTTTGAAGCCATTAAATCACTGCTCAAGCTTATGA 164
Db 127 GGAATTCATTAATCCCAATTTTGAAGCCATTAAATCACTGCTCAAGCTTATGA 186
QY 165 GCGAAGTTCGAACGAGCAAGAGTGAATTAACAAGCTCAAGCTTATGA 224
Db 187 GCGAAGTTCGAACGAGCAAGAGTGAATTAACAAGCTCAAGCTTATGA 246
QY 225 TAAGTTGATTAATGAGTCAAGAGTGTATTAAGATCAAGCTTATGA 284
Db 247 TAAGTTGATTAATGAGTCAAGAGTGTATTAAGATCAAGCTTATGA 306
QY 285 TGGCTCTCTCTTGAAGTGTATTTCTAATGATTTGGTAACTGAGAGAGCTT 344
Db 307 TGGCTCTCTCTTGAAGTGTATTTCTAATGATTTGGTAACTGAGAGAGCTT 366
QY 345 CAAGCTCTGATCACTGTTGAGAGTGGTGAAGTGTCTCTGTTGTTTCCCTGAGCT 404
Db 367 CAAGCTCTGATCACTGTTGAGAGTGGTGGTGTGCTGAGTTCCTGTTGTTCCCTGAGCT 426

QY 405 CTGTGAAGGAATTCACACTGATCCCTACCTGCAATTTCTCAGCCTTGCTTACCTGTTGC 464
Db 427 CTGTGAAGGAATTCACACTGATCCCTACCTGCAATTTCTCAGCCTTGCTTACCTGTTGC 486
QY 465 CAACATTTGGGCAACCCGAATTTCTTCCAACTTTTATCTTGGCTCCAGAGATGTCCT 524
Db 487 CAACATTTGGGCAACCCGAATTTCTTCCAACTTTTATCTTGGCTCCAGAGATGTCCT 546
QY 525 CAACAGAGCTGATACAGACAGATGGATGGTATATGTTTAAATGTCAGCTATATCTG 584
Db 547 CAACAGAGCTGATACAGACAGATGGATGGTATATGTTTAAATGTCAGCAATACCTG 606
QY 585 TCCAAAGCTGACTTATATCCCGAGTCTATTTCTGCGTGTGCTGTGAATGACAGCTT 644
Db 607 TCCAAAGCTGACTTATATCCCGAGTCTATTTCTGCGTGTGCTGTGAATGACAGCTT 666
QY 645 TTGTGAGAAATTTTGGCCGTGTTGAGCAATTCAGTATTTCAATGAAAGCAAAAGC 704
Db 667 TTGTGAGAAATTTTGGCCGTGTTGAGCAATTCAGTATTTCAATGAAAGCAAAAGC 726
QY 705 CTCGAATGATGTTGTTAGTGCACCTGTTTACTGGGATCTCCGCTCCGCAACATCCG 764
Db 727 CTCGAATGATGTTGTTAGTGCACCTGTTTACTGGGATCTCCGCTCCGCAACATCCG 786
QY 765 TATGCCCTACATCATGAAGAGATGACATGCTTTAGATGAAGCTTACAGATTTGTGA 824
Db 787 TATGCCCTACATCATGAAGAGATGACATGCTTTAGATGAAGCTTACAGATTTGTGA 846
QY 825 AGAAAAAGACCTTCTATATCTCCAACTTCATTTTCTGGGCAACTCTGCACTATGA 884
Db 847 AGAAAAAGACCTTCTATATCTCCAACTTCATTTTCTGGGCAACTCTGCACTATGA 906
QY 885 GAAAGATTTAAGACCAAGACTGAGCATGAGGCCAAGAGCAAACTCAAGCTGTGCA 944
Db 907 GAAAGATTTAAGACCAAGACTGAGCATGAGGCCAAGAGCAAACTCAAGCTGTGCA 966
QY 945 CCTGGAAGAGCAATGAACCTGCTCCCTGCTCTCAAGAGTGTGCAAGAAAGCAGAGC 1004
Db 967 CCTGGAAGAGCAATGAACCTGCTCCCTGCTCTCAAGAGTGTGCAAGAAAGCAGAGC 1026
QY 1005 GCCCTCAGTCCACCCCTGTGCGCACTCTGCTACCTCAGAGGAGCAGCAAAAGGCCGT 1064
Db 1027 GCCCTCAGTCCACCCCTGTGCGCACTCTGCTACCTCAGAGGAGCAGCAAAAGGCCGT 1086
QY 1065 GCAATCCCGCAGAGTGTGCGCAAGCTGCGCAAGCTGAGCCGTGCTTTAGAGACAGGCC 1124
Db 1087 GCAATCCCGCAGAGTGTGCGCAAGCTGCGCAAGCTGAGCCGTGCTTTAGAGACAGGCC 1146
QY 1125 GCTGTATACAGGCGCTCAGTGGGCTGCACTGTCCGAGACAGGCTGGAAGACAGCAATTA 1184
Db 1147 GCTGTATACAGGCGCTCAGTGGGCTGCACTGTCCGAGACAGGCTGGAAGACAGCAATTA 1206
QY 1185 GCTCAAGCGTTCTTCTCTGTGATATCAATCAGTTTCAATTTAGCAGCAGATGGCAGC 1244
Db 1207 GCTCAAGCGTTCTTCTCTGTGATATCAATCAGTTTCAATTTAGCAGCAGATGGCAGC 1266
QY 1245 ATCTTAATATGCTTCTCTCATCAAGAGTCTTGAATATCAAACTTCCACTAC 1304
Db 1267 ATCTTAATATGCTTCTCTCATCAAGAGTCTTGAATATCAAACTTCCACTAC 1326
QY 1305 TCTGTATGGGACCAACAGCTATGCAAGTCTCCCTTTGAGAACTATGAGAGCAGAC 1364
Db 1327 TCTGTATGGGACCAACAGCTATGCAAGTCTCCCTTTGAGAACTATGAGAGCAGAC 1386
QY 1365 TCCGGAACCGAGTCTGATTAAGAGAAAGCAAGATCCCAAGAGCTGACAGCCGAC 1424
Db 1387 TCCGGAACCGAGTCTGATTAAGAGAAAGCAAGATCCCAAGAGCTGACAGCCGAC 1446
QY 1425 GCGTTCAAGACGCAAGACAGATTTGATTTGGTGAACCAAGCAGAGCTGACAGCCG 1484
Db 1447 GCGTTCAAGACGCAAGACAGATTTGATTTGGTGAACCAAGCAGAGCTGACAGCCG 1506
QY 1485 CAGAGGTCCTTTTATCTCACTGATCGAAGTGGAGCGTGGAGCAATTAACAGAC 1544


```
QY 1201 TCTGTGGATATCAATCACTGTTTCAATATGACGACAGATGACATCTTAACTGCTTC 1260
DB 1738 TCTGTGGATATCAATCACTGTTTCAATATGACGACAGATGACATCTTAACTGCTTC 1797
QY 1261 TCTGTATGAGAAATGCTTTGGAATACTAACAACCTTCCACTACTCTGATGAGACCAAC 1320
DB 1798 TCTGTATGAGAAATGCTTTGGAATACTAACAACCTTCCACTACTCTGATGAGACCAAC 1857
QY 1321 AAGCTATGCGAGTTCCTCCCTGTGTCAGAACTATGAGACAGACTCCGAAACCAAGTCT 1380
DB 1858 AAGCTATGCGAGTTCCTCCCTGTGTCAGAACTATGAGACAGACTCCGAAACCAAGTCT 1917
QY 1381 GATAGAGAGAAAGCAGATCCCAAGAGCTGACAGCCGACGCTTTCAGACAGCAG 1440
DB 1918 GATAGAGAGAAAGCAGATCCCAAGAGCTGACAGCCGACGCTTTCAGACAGCAG 1977
QY 1441 AACAAGCCATGATTCGGTGCAGAACGACGACAGTGGCAACGACGCTTCCTTTTGA 1500
DB 1978 AACAAGCCATGATTCGGTGCAGAACGACGACAGTGGCAACGACGCTTCCTTTTGA 2037
QY 1501 TCTCCACTGATCGAAGTGGAGAGCTGGAGCAATTCACACACAGCTTCCTTTTGGC 1560
DB 2038 TCTCCACTGATCGAAGTGGAGAGCTGGAGCAATTCACACACAGCTTCCTTTTGGC 2097
QY 1561 CTTTCCACACAGCCAGACAGACCTTCAAGAGTCTGCTGGCTGGCTTTAAAGGCTGGCAC 1620
DB 2098 CTTTCCACACAGCCAGACAGACCTTCAAGAGTCTGCTGGCTGGCTTTAAAGGCTGGCAC 2157
QY 1621 TCGGATATCTTGGCCCCCAGACCTTACCCCTTCCCTGACGACAGCTGTATTTTGGC 1680
DB 2158 TCGGATATCTTGGCCCCCAGACCTTACCCCTTCCCTGACGACAGCTGTATTTTGGC 2217
QY 1681 ACAAGTCTCTCACTTCTACTGCTGCTGACGACATCTTACGAGAGAGCTGACATTTACT 1740
DB 2218 ACAAGTCTCTCACTTCTACTGCTGCTGACGACATCTTACGAGAGAGCTGACATTTACT 2277
QY 1741 GCCTACAGCTGSCAGCCAGCTGCCCCACTTTGGGGAGAGCCAAATCTATTTCTGTGGCAGCGG 1800
DB 2278 GCCTACAGCTGSCAGCCAGCTGCCCCACTTTGGGGAGAGCCAAATCTATTTCTGTGGCAGCGG 2337
QY 1801 CAGAGCCAAAGTGAAGAGCTGTAATGCGGCGAGAGCTGCGATGAAGAGAGCCCTTTTGA 1860
DB 2338 CAGAGCCAAAGTGAAGAGCTGTAATGCGGCGAGAGCTGCGATGAAGAGAGCCCTTTTGA 2397
QY 1861 AAGCATTTTAAAGCGAGAGCTGCGCAATTTGGAATTTGAGAGCATCTGCAAGAAC 1920
DB 2398 AAGCATTTTAAAGCGAGAGCTGCGCAATTTGGAATTTGAGAGCATCTGCAAGAAC 2457
QY 1921 AGGTCAAGGGAAGAGCTGGGGAAGTGGCAAGTCAAGCTTTTGGGGGAGCATTTGGA 1980
DB 2458 AGGTCAAGGGAAGAGCTGGGGAAGTGGCAAGTCAAGCTTTTGGGGGAGCATTTGGA 2517
QY 1981 ATCATTTGAGGTCTCC 1995
DB 2518 ATCATTTGAGGTCTCC 2532
```

RESULT 12

```
US-10-094-749-673
; Sequence 673, Application US/10094749
; Publication No. US20030219741A1
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
```

```
APPLICANT: IRIE, RYOTARO
APPLICANT: TAMECHIKA, ICHIRO
APPLICANT: SEKI, NAOHICO
APPLICANT: YOSHIDAKA, TSUTOMU
APPLICANT: OTSUKA, MOTOYUKI
APPLICANT: NAGAHARI, KENJI
APPLICANT: MASUHO, YASUHIKO
TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
FILE REFERENCE: 08435/0160
CURRENT APPLICATION NUMBER: US/10/094,749
PRIOR FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: 60/350,435
PRIOR FILING DATE: 2002-01-24
PRIOR APPLICATION NUMBER: JP 2001-328381
NUMBER OF SEQ ID NOS: 3381
SOFTWARE: Patencin Ver. 2.1
SEQ ID NO: 673
LENGTH: 2102
TYPE: DNA
ORGANISM: Homo sapiens
US-10-094-749-673
```

```
Query Match 89.8%; Score 1791; DB 16; Length 2102;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1991; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 1 ATGCCCATGAGATGATGGAACCTCAATTGTTTACTGAGAGTGTGCTGCTGGAA 60
DB 56 ATGCCCATGAGATGATGGAACCTCAATTGTTTACTGAGAGTGTGCTGCTGGAA 115
QY 61 AGTGAACGAAAAAGTCTGCTAATTGATGAGCGCGCATTTTGTGAATCAATATCC 120
DB 116 AGTGAACGAAAAAGTCTGCTAATTGATGAGCGCGCATTTTGTGAATCAATATCC 175
QY 121 CACATTTTGAAGCCATTAATATCACTGCTCAAGCTTATGAAGGAGTTGCAACG 180
DB 176 CACATTTTGAAGCCATTAATATCACTGCTCAAGCTTATGAAGGAGTTGCAACG 235
QY 181 GACAAAGTTAATTAAGAGCTCATCCAGATTGACGAAACATTAAGTTGACATTGAT 240
DB 236 GACAAAGTTAATTAAGAGCTCATCCAGATTGACGAAACATTAAGTTGACATTGAT 295
QY 241 TGACGTGAGAGGTTGATGTTTACATCAAGCTCCCAAGATGTTGCTCTCTTTGA 300
DB 296 TGACGTGAGAGGTTGATGTTTACATCAAGCTCCCAAGATGTTGCTCTCTTTGA 355
QY 301 GACTGTTTCTCATCTGACTTCTGGGTAATGAGAGAGCTTCAACTCTGTACCTG 360
DB 356 GACTGTTTCTCATCTGACTTCTGGGTAATGAGAGAGCTTCAACTCTGTACCTG 415
QY 361 CTTCAGAGTGGTTCCTGAGTTCCTGCTGTTTCCCTGGCTGTGGAAGAAATCC 420
DB 416 CTTCAGAGTGGTTCCTGAGTTCCTGCTGTTTCCCTGGCTGTGGAAGAAATCC 475
QY 421 ACTTAGCTCTACCTGACTTCTCAGCCTGCTTACCTGTTCCAAATTTGGGCAAC 480
DB 476 ACTTAGCTCTACCTGACTTCTCAGCCTGCTTACCTGTTCCAAATTTGGGCAAC 535
QY 481 CGAATTTTCCCAATCTTATCTTGGCTGCGAGCGAGATGTCCTCAACAGACTGATG 540
DB 536 CGAATTTTCCCAATCTTATCTTGGCTGCGAGCGAGATGTCCTCAACAGACTGATG 595
QY 541 CAGCAAAATGGATGTTGTTATGTTAATGCCAGCTTATCTGCTCCAAAGCTGACTT 600
DB 596 CAGCAAAATGGATGTTGTTATGTTAATGCCAGCAATACCTGCTCCAAAGCTGACTT 655
QY 601 ATCCCGAGTCTCATTTCTGCTGCTGCTGCTGATGACAGCTTTTGTGAAGAAATTTG 660
DB 656 ATCCCGAGTCTCATTTCTGCTGCTGCTGCTGATGACAGCTTTTGTGAAGAAATTTG 715
QY 661 CCGTGTGGAACAATGATGATTTTCAATTGAGAGCAAAAGCTTCAATGATGTGT 720
```

D 716 CCGTGGTGGACAAATCAGATATTCAATGAGAAAGCAAAAGCTCAATGATGTGTT 775
Q 721 CTAAGTCACTGTTTACTGGGATCTCCCGTCCGCCAATCGCTATGGCTTACATCAG 780
D 776 CTAGTGCACGTGTACTGGGATCTCCCGTCCGCCAATCGCTATGGCTTACATCAG 835
Q 781 AAGAGATGGAACATGCTTTAGATGAGCTTACAGATTTGTGAAAAGAAAGACCTACT 840
D 836 AAGAGATGGAACATGCTTTAGATGAGCTTACAGATTTGTGAAAAGAAAGACCTACT 895
Q 841 ATATCTCAAAATCTCAATTTCTGGGCCAACTCTGAGATGAGAGAAATTAAGAAC 900
D 896 ATATCTCAAAATCTCAATTTCTGGGCCAACTCTGAGATGAGAGAAATTAAGAAC 955
Q 901 CAGACTGGAGCATCAGGGCCAAAGAGCAAACTCAGCTGTGACCTGGAGAAAGCAAA 960
D 956 CAGACTGGAGCATCAGGGCCAAAGAGCAAACTCAGCTGTGACCTGGAGAAAGCAAA 1015
Q 961 GAACTGTCCCTGCTGCTCAGAGGGTGGACAGAAAAGCGAGAGCCCTCAGTCCACCC 1020
D 1016 GAACTGTCCCTGCTGCTCAGAGGGTGGACAGAAAAGCGAGAGCCCTCAGTCCACCC 1075
Q 1021 TGTGCCGACTGTGCTACCTCAGAGGAGCAGAGCAAAAGGCCGTGATCCCGCAGCGTG 1080
D 1076 TGTGCCGACTGTGCTACCTCAGAGGAGCAGAGCAAAAGGCCGTGATCCCGCAGCGTG 1135
Q 1081 CCCAGCGTGGCCAGCGCTGACCGCTGTGAGAGACAGCCCGCTGGTACAGGCGCTC 1140
D 1136 CCCAGCGTGGCCAGCGCTGACCGCTGTGAGAGACAGCCCGCTGGTACAGGCGCTC 1195
Q 1141 AGTGGGCTGACCTGTCCGAGACAGGCTGAGAAAGACAGCAATTAAGCTCAAGGCTTC 1200
D 1196 AGTGGGCTGACCTGTCCGAGACAGGCTGAGAAAGACAGCAATTAAGCTCAAGGCTTC 1255
Q 1201 TCTCTGATATCAAAATCAGTTTCATATTCAGCCAGCATGAGCATCTTACATGCTTC 1260
D 1256 TCTCTGATATCAAAATCAGTTTCATATTCAGCCAGCATGAGCATCTTACATGCTTC 1315
Q 1261 TCTCTCATCAGAGATGCTTTGGAATCTAACAACTTCCATCTGATGGAGCAAC 1320
D 1316 TCTCTCATCAGAGATGCTTTGGAATCTAACAACTTCCATCTGATGGAGCAAC 1375
Q 1321 AAGCTATGCGAGTTCTCCCTGTTCAAGAACTATCGAGAGCATCCCGAAACAGTCTC 1380
D 1376 AAGCTATGCGAGTTCTCCCTGTTCAAGAACTATCGAGAGCATCCCGAAACAGTCTC 1435
Q 1381 GATTAAGAGGAAAGCCAGCATCCCAAGAGCTGACAGCGCCAGCTTCAGACAGCAG 1440
D 1436 GATTAAGAGGAAAGCCAGCATCCCAAGAGCTGACAGCGCCAGCTTCAGACAGCAG 1495
Q 1441 AGCAGCGATTGATCGGTGAGAACCAAGAGGTCGACCGCCCAAGAGTCCCTTTTA 1500
D 1496 AGCAGCGATTGATCGGTGAGAACCAAGAGGTCGACCGCCCAAGAGTCCCTTTTA 1555
Q 1501 TCTCCACTGATCGAGTGGAGCGCTGAGAGCAATTAACCAACAGCTTCCTTTGCGC 1560
D 1556 TCTCCACTGATCGAGTGGAGCGCTGAGAGCAATTAACCAACAGCTTCCTTTGCGC 1615
Q 1561 CTTTCCAGCAGCAGAGCATCTCAGAAAGTGTGAGCTGGGCTTTAAGGGCTGGCAC 1620
D 1616 CTTTCCAGCAGCAGAGCATCTCAGAAAGTGTGAGCTGGGCTTTAAGGGCTGGCAC 1675
Q 1621 TCGGATATCTTGGCCGCCAGACCTTACCCCTTCCCTGACAGAGCTGTATTTTGGC 1680
D 1676 TCGGATATCTTGGCCGCCAGACCTTACCCCTTCCCTGACAGAGCTGTATTTTGGC 1735
Q 1681 ACAGAGTCTCAGACTTCTACTGCTCAGAGCATCTACGAGAGGCAAGTGCAGTTACTCT 1740
D 1736 ACAGAGTCTCAGACTTCTACTGCTCAGAGCATCTACGAGAGGCAAGTGCAGTTACTCT 1795
Q 1741 GCCTAAGCTGACAGCAGCTGACCTGCTGAGAGCAAAAGTCTATCTGTGCGAGGCGG 1800
D 1796 GCCTAAGCTGACAGCAGCTGACCTGCTGAGAGCAAAAGTCTATCTGTGCGAGGCGG 1855

Q 1801 CAGAAAGCAAGTACAGAGCTGACTCGGGCGGAGAGCTGGCATGAGAGAGCCCTTTGAA 1860
D 1856 CAGAAAGCAAGTACAGAGCTGACTCGGGCGGAGAGCTGGCATGAGAGAGCCCTTTGAA 1915
Q 1861 AAGCAGTTTAAAGCAGAGAGCTGCCAAATGGAATTTGAGAGAGCATGTCAGAGAAC 1920
D 1916 AAGCAGTTTAAAGCAGAGAGCTGCCAAATGGAATTTGAGAGAGCATGTCAGAGAAC 1975
Q 1921 AGGTACCGGAGAGAGCTGGGGAAGTGGGCACTCAGTCTTCTGAGGAGCATGAA 1980
D 1976 AGGTACCGGAGAGAGCTGGGGAAGTGGGCACTCAGTCTTCTGAGGAGCATGAA 2035
Q 1981 ATCATTGAGTCTCC 1995
D 2036 ATCATTGAGTCTCC 2050

Search completed: June 22, 2004, 16:36:33
Job time : 1178 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 22, 2004, 09:24:40 ; Search time 172 Seconds

(without alignments)
6436.785 Million cell updates/sec

Title: US-10-029-345A-108_COPY_538_2532

Perfect score: 1995
Sequence: 1 atggcccatgagatgagtgatg.....tggaatcatgagtgctcc 1995

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 682709 seqs, 277475446 residues

Word size : 1419

Total number of hits satisfying chosen parameters: 2

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 75 summaries

Database : Issued_Patents_NA:*

1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/prodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/prodata/2/ina/PCUS_COMB.seq:*
6: /cgn2_6/prodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1893	94.9	1998	4 US-09-816-494-3	Sequence 3, Appl1
2	1893	94.9	3544	4 US-09-816-494-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-09-816-494-3
; Sequence 3, Application US/09816494
; Patent No. 6664089
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel A.
; TITLE OF INVENTION: 38692 AND 21117, NOVEL DUAL SPECIFICITY
; TITLE OF INVENTION: PHOSPHATASE MOLECULES AND USES THEREFOR
; FILE REFERENCE: 10448-030002
; CURRENT APPLICATION NUMBER: US/09/816,494
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 60/191,858
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1998
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-816-494-3

Query Match	94.9%;	Score 1893;	DB 4;	Length 1998;
Best Local Similarity	99.9%;	Pred. 0;		
Matches 1993;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;
QY	1	ATGGCCCATGAGATGATGGAACTCAATTTGTTACTGAGAGTGGTGCTGCTGGAA	60	
DB	1	ATGGCCCATGAGATGATGGAACTCAATTTGTTACTGAGAGTGGTGCTGCTGGAA	60	
QY	61	AGTGAACGGAAAAAGTGTCTGCTAATTGATAGCCGGCCATTGTGGAAATCAATACATCC	120	
DB	61	AGTGAACGGAAAAAGTGTCTGCTAATTGATAGCCGGCCATTGTGGAAATCAATACATCC	120	
QY	121	CACATTTTGGAGGCAATTAATATCACTGCTCCAGCTTATGAAGGAAAGTTGCAACAG	180	
DB	121	CACATTTTGGAGGCAATTAATATCACTGCTCCAGCTTATGAAGGAAAGTTGCAACAG	180	
QY	181	GACAAAGTTTAATTAACAGAGCTCATCCAGCATTCAGCGAAACATAAGTTGACATTGAT	240	
DB	181	GACAAAGTTTAATTAACAGAGCTCATCCAGCATTCAGCGAAACATAAGTTGACATTGAT	240	
QY	241	TGCAGTCAAGAGTTGATTTAGATCAAGTCCCAAGATTTGCTCTCTCTTCA	300	
DB	241	TGCAGTCAAGAGTTGATTTAGATCAAGTCCCAAGATTTGCTCTCTCTTCA	300	
QY	301	GACTTTTCTCACTGATCTTGCGGTAACCTGGAAGAGCTTCAACTGTTCACTG	360	
DB	301	GACTTTTCTCACTGATCTTGCGGTAACCTGGAAGAGCTTCAACTGTTCACTG	360	
QY	361	CTTGACAGTGGGTTTCTGATCTCTGTTGTTCTTGCGCTCTGTGAGGAAATCC	420	
DB	361	CTTGACAGTGGGTTTCTGATCTCTGTTGTTCTTGCGCTCTGTGAGGAAATCC	420	
QY	421	ACTTATGCTTCACTGATCTTGCGGTAACCTGGAAGAGCTTCAACTGTTCACTG	480	
DB	421	ACTTATGCTTCACTGATCTTGCGGTAACCTGGAAGAGCTTCAACTGTTCACTG	480	
QY	481	CGAATCTTCCCAATCTTATCTTGCTGCGAGGAGTCTCTCAACAGAGCTGATA	540	
DB	481	CGAATCTTCCCAATCTTATCTTGCTGCGAGGAGTCTCTCAACAGAGCTGATA	540	
QY	541	CAGCAAAATGAGTGGTATTTGTTAAATCCAGCTATACCTGTCCAAAGCTGACTTT	600	
DB	541	CAGCAAAATGAGTGGTATTTGTTAAATCCAGCTATACCTGTCCAAAGCTGACTTT	600	
QY	601	ATCCCGAGTCTCATTTCTGCTGCTGCTGCTGATGACAGCTTTGTGAGAAATTTTG	660	
DB	601	ATCCCGAGTCTCATTTCTGCTGCTGCTGCTGATGACAGCTTTGTGAGAAATTTTG	660	
QY	661	CCGTGTTGACAAATCAGTATGATTTCAATGAGAAAGCAAAAGCTTCAATGATGATGTT	720	
DB	661	CCGTGTTGACAAATCAGTATGATTTCAATGAGAAAGCAAAAGCTTCAATGATGATGTT	720	
QY	721	CTAGTCACTGTTTACCTGGATCTCCGCTCCGACCATGCTATCCGCTACATCATG	780	
DB	721	CTAGTCACTGTTTACCTGGATCTCCGCTCCGACCATGCTATCCGCTACATCATG	780	
QY	781	AAGAGATGACATGCTTTAGATGAGCTTACAGATTTGTGAGAAAGAAAGAAAGCTTACT	840	
DB	781	AAGAGATGACATGCTTTAGATGAGCTTACAGATTTGTGAGAAAGAAAGAAAGCTTACT	840	
QY	841	ATATCTCCAAATTTTCTGCGCAATCTTGGACCTATGATGAGAAAGATTAAGAAC	900	
DB	841	ATATCTCCAAATTTTCTGCGCAATCTTGGACCTATGATGAGAAAGATTAAGAAC	900	
QY	901	CAGACTGAGACATCAGGCGCAAAAGCAAACTCAAGCTCTGACCTGAGAGGCAAAAT	960	
DB	901	CAGACTGAGACATCAGGCGCAAAAGCAAACTCAAGCTCTGACCTGAGAGGCAAAAT	960	
QY	961	GAACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1020	
DB	961	GAACTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1020	

```

/ GENERAL INFORMATION:
/ APPLICANT: Meyers, Rachel A.
/ TITLE OF INVENTION: 38692 AND 21117, NOVEL DUAL SPECIFICITY
/ TITLE OF INVENTION: PHOSPHATASE MOLECULES AND USES THEREFOR
/ FILE REFERENCE: 10448-030002
/ CURRENT APPLICATION NUMBER: US/09/816,494
/ PRIOR FILING DATE: 2001-03-23
/ PRIOR APPLICATION NUMBER: US 60/191,858
/ PRIORITY DATE: 2000-03-24
/ NUMBER OF SEQ ID NOS: 10
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 1
/ LENGTH: 3544
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (589)...(2583)
/ US-09-816-494-1

Query Match          94.9%; Score 1893; DB 4; Length 3544;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1993; Conservative 0; Mismatches 2; Indels 0; Gaps 0

QY      1 ATGGCCCATGAGTGAATGGAACTCAAAATTTGTTACTGAGAGGTTGGTGGCTCTGCTGGAA 60
Db      589 ATGGCCCATGAGTGAATGGAACTCAAAATTTGTTACTGAGAGGTTGGTGGCTCTGCTGGAA 648
QY      61 AATGGAACGAAAAAGTCTGCTAATTTGATGATGACCGGCAATTTGTGGAATACATATCATCC 120
Db      649 AATGGAACGAAAAAGTCTGCTAATTTGATGATGACCGGCAATTTGTGGAATACATATCATCC 708
QY      121 CACATTTTGGAAACCATTAATATATCACTCTCAAGCTTATTAAGCGAAAGTTGCAACAG 180
Db      709 CACATTTTGGAAACCATTAATATATCACTCTCAAGCTTATTAAGCGAAAGTTGCAACAG 768
QY      181 GACAAAGTTTAATTAACAGAGCTCATCCAGCATTCAGGAAACATTAAGTTGACATTTGAT 240
Db      769 GACAAAGTTTAATTAACAGAGCTCATCCAGCATTCAGGAAACATTAAGTTGACATTTGAT 828
QY      241 TGCAGTCAGAAAGTTGTGTTTACGATCAAAAGCTCCCAAGTGTGGCTCTCTCTCTTCA 300
Db      829 TGCAGTCAGAAAGTTGTGTTTACGATCAAAAGCTCCCAAGTGTGGCTCTCTCTCTTCA 888
QY      301 GACTGTTTCTCACTGACTCTCTGGGTAACTGAGAGAGACTTCAACTCTGTTCACCTG 360
Db      889 GACTGTTTCTCACTGACTCTCTGGGTAACTGAGAGAGACTTCAACTCTGTTCACCTG 948
QY      361 CTTCGAGGTGGTTGCTGAGTTCTCTGTTGTTTCCCTGGCCCTCTGTGAAGAAATTC 420
Db      949 CTTCGAGGTGGTTGCTGAGTTCTCTGTTGTTTCCCTGGCCCTCTGTGAAGAAATTC 1008
QY      421 ACTTAAGTCCCTAATCTGATTTCTCAAGCTTGTACTTGTGCAACATTTGGGCCAAC 480
Db      1009 ACTTAAGTCCCTAATCTGATTTCTCAAGCTTGTACTTGTGCAACATTTGGGCCAAC 1068
QY      481 CGAATTTCTTCCCAATCTTTATCTTGGGTGCGCAGCGAGATGTCCTCAACAGAGAGCTGATA 540
Db      1069 CGAATTTCTTCCCAATCTTTATCTTGGGTGCGCAGCGAGATGTCCTCAACAGAGAGCTGATA 1128
QY      541 CAGAGAAATGGGATTTGGTTATGTTTAATGCAAGCTATATCTGTCCAAAGCTGTGACTTT 600
Db      1129 CAGAGAAATGGGATTTGGTTATGTTTAATGCAAGCAATATCTGTCCAAAGCTGTGACTTT 1188
QY      601 ATCCCAAGTCTCAATTTCTGCGGTTGCGCTGTGAATGACAGCTTTGTGAAGAAATTTTG 660

```

Query Match	Best Local Similarity	94.9%;	Score 1893;	DB 4;	Length 3544;
Matches 1993;	Conservative	0;	Mismatches 2;	Indels 0;	Gaps 0
QY	1	ATGCCCATATGAGATGATTGGAACTCCAAATTTGTTACTGAGAGAGTTGGTGGCTCTGCTGGAA	60		
Db	589	ATGGCCCATATGATGATTTGGAACTCCAAATTTGTTACTGAGAGAGTTGGTGGCTCTGCTGGAA	648		
QY	61	AGTGGAAACGAAAAAAGTCTGCTAATTGGATTGATAGCCGGCCATTGTGGATATACATATCATCC	120		
Db	649	AGTGGAAACGAAAAAAGTCTGCTAATTGGATTGATAGCCGGCCATTGTGGATATACATATCATCC	708		
QY	121	CACATTTTGGGAACCCCTTAATATCAACTCTCCAAAGCTTATGAAACGAAGTTGGCAACAG	180		
Db	709	CACATTTTGGGAACCCCTTAATATCAACTCTCCAAAGCTTATGAAACGAAGTTGGCAACAG	768		
QY	181	GACAAAGTGTAAATTACAGAGCTCAATCCAGATTTAGGGAACATAAGSTTGAACATTGAT	240		
Db	769	GACAAAGTGTAAATTACAGAGCTCAATCCAGATTTAGGGAACATAAGSTTGAACATTGAT	828		
QY	241	TGCAGTCAGAAAGTTGATTTAGCATCAAAAGCTCCCAAGATGTTGCTCTCTCTCTTCA	300		
Db	829	TGCAGTCAGAAAGTTGATTTAGCATCAAAAGCTCCCAAGATGTTGCTCTCTCTCTTCA	888		
QY	301	GACTGTTTTCTACAGTACTCTCTGGGTAAACTGGAAGAAAGCTTCAACTCTGTTCACTG	360		
Db	889	GACTGTTTTCTACAGTACTCTCTGGGTAAACTGGAAGAAAGCTTCAACTCTGTTCACTG	948		
QY	361	CTTGCAGGTGGGTTTGCTGAGTCTCTCGTGTGTTCCCTGGCCCTCTGGAAGGAAATATCC	420		
Db	949	CTTGCAGGTGGGTTTGCTGAGTCTCTCGTGTGTTCCCTGGCCCTCTGGAAGGAAATATCC	1008		
QY	421	ACTCTAGTCCCTACCTGCAATTTCTCAGCCCTGCTTACCTGTGGCAAACTTGGGCCAAC	480		
Db	1009	ACTCTAGTCCCTACCTGCAATTTCTCAGCCCTGCTTACCTGTGGCAAACTTGGGCCAAC	1068		
QY	481	CGAATTCTTCCCAATCTTTATCTTGGGTGGCCAGCGAGATGCTCTCAACAGAGCTGATA	540		
Db	1069	CGAATTCTTCCCAATCTTTATCTTGGGTGGCCAGCGAGATGCTCTCTCAACAGAGCTGATA	1128		
QY	541	CAGCAGATGGGATTTGTTATGTGTTAAATGCGCAGCTATACCTGTCCAAAGCCTGACTTT	600		
Db	1129	CAGCAGATGGGATTTGTTATGTGTTAAATGCGCAGCTATACCTGTCCAAAGCCTGACTTT	1188		
QY	601	ATCCCCCAAGTCTCAATTTCTGCGGTGGCCGTGGAATGACAGCTTTTGTGAGAAAATTTTG	660		
Db	1189	ATCCCCCAAGTCTCAATTTCTGCGGTGGCCGTGGAATGACAGCTTTTGTGAGAAAATTTTG	1248		
QY	661	CCGCTGTTGGAACAATCAGTAGATTTCAATTGGAAAGCAAAAAGCCTCCATGATGATGTT	720		
Db	1249	CCGCTGTTGGAACAATCAGTAGATTTCAATTGGAAAGCAAAAAGCCTCCATGATGATGTT	1308		
QY	721	CTAGTCACTGTTTACCTGGGATCTCCGCTCCGCAACATGAGTATAGGCTCATCATATAG	780		

Db 1309 CTAGTGCATGTTAGCTGGGATCTCCGCTCCGCCACATCGCTATCGCTCATCATG 1368
 Qy 781 AAGAGGATGGAACATGCTTTTATGATGAAGCTTACAGATTTGTGAAGAAAAAAGACTACT 840
 Db 1369 AAGAGGATGGAACATGCTTTTATGATGAAGCTTACAGATTTGTGAAGAAAAAAGACTACT 1428
 Qy 841 ATATCTCAAACTTCAATTTTCTGGGCAACTCTCTGACTATGAGAAAGATTTAAGAC 900
 Db 1429 ATATCTCAAACTTCAATTTTCTGGGCAACTCTCTGACTATGAGAAAGATTTAAGAC 1488
 Qy 901 CAGACTGAGACATCAGGCGCAAGAGCAAACTCAAGCTGTGCACTGAGAACCAAT 960
 Db 1489 CAGACTGAGACATCAGGCGCAAGAGCAAACTCAAGCTGTGCACTGAGAACCAAT 1548
 Qy 961 GAACTGTCTCTGTCTTCAAGAGGTGACAGAGAAAGCAAGCCCTCAGTCCACC 1020
 Db 1549 GAACTGTCTCTGTCTTCAAGAGGTGACAGAGAAAGCAAGCCCTCAGTCCACC 1608
 Qy 1021 TGTGCGGACTGTGCTCTGAGAGGCGAGAGCAAAAGCCGCTGCACTCCGCGCAGCGTG 1080
 Db 1609 TGTGCGGACTGTGCTCTGAGAGGCGAGAGCAAAAGCCGCTGCACTCCGCGCAGCGTG 1668
 Qy 1081 CCCAGCGTGCAGCGCTGAGCGCTGCTGTAGAGGACAGCCGCTGTATCAGGCGCTC 1140
 Db 1669 CCCAGCGTGCAGCGCTGAGCGCTGCTGTAGAGGACAGCCGCTGTATCAGGCGCTC 1728
 Qy 1141 AGTGGGCTGCACTGTCTCCGAGACAGGCTGGAAGACAGCAATTAAGCTCAAGCTTCTTC 1200
 Db 1729 AGTGGGCTGCACTGTCTCCGAGACAGGCTGGAAGACAGCAATTAAGCTCAAGCTTCTTC 1788
 Qy 1201 TCTCTGATATCAATATGATTTATATTCAGCCAGCATGAGCATCTTACATGCGTTC 1260
 Db 1789 TCTCTGATATCAATATGATTTATATTCAGCCAGCATGAGCATCTTACATGCGTTC 1848
 Qy 1261 TCTCTATCAGAAAGATGCTTGAATACTAATAAATCTTCACTACTGATGGAGCAAC 1320
 Db 1849 TCTCTATCAGAAAGATGCTTGAATACTAATAAATCTTCACTACTGATGGAGCAAC 1908
 Qy 1321 AAGCTATGCGAGTTCTCCCTGTTTCAAGAACTATCGGAGCAGACTCCGAAACAGTCTT 1380
 Db 1909 AAGCTATGCGAGTTCTCCCTGTTTCAAGAACTATCGGAGCAGACTCCGAAACAGTCTT 1968
 Qy 1381 GATTAAGAGAGAGCCAGCATCTCCCAAGAACTGCAACCGCCAGGCTTTCAGACAGCCAG 1440
 Db 1969 GATTAAGAGAGAGCCAGCATCTCCCAAGAACTGCAACCGCCAGGCTTTCAGACAGCCAG 2028
 Qy 1441 AGCAAGGATTTGATTTGCTGATCAGAACAGAGCAGTGGACCGCCAGAGGTCCCTTTTA 1500
 Db 2029 AGCAAGGATTTGATTTGCTGATCAGAACAGAGCAGTGGACCGCCAGAGGTCCCTTTTA 2088
 Qy 1501 TCTCCACTGATGGAATGGAAGGTGAGAGCAATTAACCAACAGCTTCTTTTGGGC 1560
 Db 2089 TCTCCACTGATGGAATGGAAGGTGAGAGCAATTAACCAACAGCTTCTTTTGGGC 2148
 Qy 1561 CTTTCCACGAGCAGCAGCACTTCAAGAACTGTGCTGAGGCTTAAAGGCTGCGAC 1620
 Db 2149 CTTTCCACGAGCAGCAGCACTTCAAGAACTGTGCTGAGGCTTAAAGGCTGCGAC 2208
 Qy 1621 TCGGATTTCTTGGGCCCCCAGAGCTTACCCCTTCCCTGACAGCAGCTGTATTTTGGC 1680
 Db 2209 TCGGATTTCTTGGGCCCCCAGAGCTTACCCCTTCCCTGACAGCAGCTGTATTTTGGC 2268
 Qy 1681 ACAGAGTCTTCAACTTCTACTGTCTGACCACTTACAGGAGGCAAGTTCAGTACTCT 1740
 Db 2269 ACAGAGTCTTCAACTTCTACTGTCTGACCACTTACAGGAGGCAAGTTCAGTACTCT 2328
 Qy 1741 GCTTACAGCTGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1800
 Db 2329 GCTTACAGCTGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2388
 Qy 1801 CAGAAAGCAAGTGAAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1860

Db 2389 CAGAAAGCAAGTGAAGAGCTGACTGCGGCGGAGCTGCTGATGAAGAGCCCTTTGAA 2448
 Qy 1861 AAGCAATTTAAACGCAAGAGCTGCGCAATTTGAGAGAGCATCAATGTCAGAGAAC 1920
 Db 2449 AAGCAATTTAAACGCAAGAGCTGCGCAATTTGAGAGAGCATCAATGTCAGAGAAC 2508
 Qy 1921 AGCTCAGGGAAGAGCTGCGGAAAGTGGCACTGACTTATGCTTTTGGGCAAGCATGAA 1980
 Db 2509 AGCTCAGGGAAGAGCTGCGGAAAGTGGCACTGACTTATGCTTTTGGGCAAGCATGAA 2568
 Qy 1981 ATCATTGAGGTCTCC 1995
 Db 2569 ATCATTGAGGTCTCC 2583

Search completed: June 22, 2004, 16:16:35
 Job time : 173 secs

1995
 -575
 1340



This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 13:25:03 ; Search time 21 Seconds
(without alignments)
3046.065 Million cell updates/sec

Title: US-10-029-345A-109

Perfect score: 665

Sequence: 1 MAHEMIGTQIVTERLVALL.....LGKVSQSSFSGSMELIEVS 665

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0

Total number of hits satisfying chosen parameters: 57676

Minimum DB seq length: 473

Maximum DB seq length: 200000000

Post-Processing: Listing first 75 summaries

Database :

1: PIR78:*
2: pir1:*
3: pir2:*
4: pir3:*
5: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9	1.4	619	2	T15969
2	8	1.2	727	2	AB2353
3	8	1.2	1308	2	T15280
4	7	1.1	475	2	B70143
5	7	1.1	478	2	S17836
6	7	1.1	479	2	T47415
7	7	1.1	488	2	AE0001
8	7	1.1	507	2	G84577
9	7	1.1	509	2	D86911
10	7	1.1	509	2	T10013
11	7	1.1	510	2	A38337
12	7	1.1	511	2	AE3405
13	7	1.1	515	2	S61889
14	7	1.1	524	2	B97864
15	7	1.1	535	2	S40461
16	7	1.1	535	2	S40462
17	7	1.1	546	2	C72453
18	7	1.1	548	2	G81959
19	7	1.1	564	2	T72268
20	7	1.1	566	2	S50979
21	7	1.1	579	2	T10494
22	7	1.1	583	2	T11624
23	7	1.1	585	2	S55205
24	7	1.1	596	2	T17333
25	7	1.1	601	2	T13054
26	7	1.1	615	2	G90575
27	7	1.1	622	2	B86431
28	7	1.1	628	1	A39262
29	7	1.1	628	1	A33333

30	7	1.1	634	2	G90779	hypothetical prote
31	7	1.1	657	2	S32991	hypothetical prote
32	7	1.1	657	2	F81896	probable phage tra
33	7	1.1	661	2	A81125	bacteriophage tran
34	7	1.1	661	2	S45131	probable membrane
35	7	1.1	674	2	C81505	hypothetical prote
36	7	1.1	674	2	E72033	hypothetical prote
37	7	1.1	680	2	B65590	hypothetical prote
38	7	1.1	680	2	G82618	pilus biogenesis p
39	7	1.1	693	2	S01991	transforming prote
40	7	1.1	700	1	S33704	transforming prote
41	7	1.1	704	2	T38117	probable protein k
42	7	1.1	731	2	T17002	probable beta-gala
43	7	1.1	745	2	C99995	hypothetical prote
44	7	1.1	745	2	H85840	hypothetical prote
45	7	1.1	757	2	E82013	probable transcrip
46	7	1.1	757	2	A81242	transcription acce
47	7	1.1	778	2	E97224	ATP-dependent lon
48	7	1.1	784	2	T51759	glutamate-ammonia
49	7	1.1	795	2	T34468	hypothetical prote
50	7	1.1	828	2	T03544	hypothetical prote
51	7	1.1	858	2	UC7683	laase receptor TIR
52	7	1.1	866	2	F88481	protein C16A3.1 f1
53	7	1.1	896	2	D96556	hypothetical prote
54	7	1.1	930	2	T00403	hypothetical prote
55	7	1.1	938	2	S20480	trac protein - Esc
56	7	1.1	946	2	T38100	rho-GTPase-activat
57	7	1.1	966	2	G70838	probable mmp11 pr
58	7	1.1	1052	2	T04439	hypothetical prote
59	7	1.1	1071	2	S48378	probable membrane
60	7	1.1	1099	2	A55405	adenylate cyclase
61	7	1.1	1115	2	T29012	hypothetical prote
62	7	1.1	1136	2	F96564	hypothetical prote
63	7	1.1	1165	2	D59433	C. elegans protein
64	7	1.1	1234	2	T00363	hypothetical prote
65	7	1.1	1264	2	B84534	probable retroelem
66	7	1.1	1277	2	T14152	synaptic scaffold
67	7	1.1	1309	2	S62457	hypothetical prote
68	7	1.1	1318	2	T38568	hypothetical prote
69	7	1.1	1454	2	E84535	probable retroelem
70	7	1.1	1461	2	E84589	probable retroelem
71	7	1.1	1465	2	T30891	PHY3 protein - mal
72	7	1.1	1642	2	I51018	cobra venom factor
73	7	1.1	1737	2	T00209	MEGR8 protein - hu
74	7	1.1	1792	2	T08878	supervillin P205 -
75	7	1.1	2077	1	WZBZ24	240k tegument prot

ALIGNMENTS

RESULT 1
T15969
hypothetical protein F08B1.1 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C/Accession: T15969
R/Chisoe: S.
submitted to the EMBL Data Library, July 1995
A/Description: The sequence of C. elegans consltd F08B1.
A/Reference number: Z18439
A/Accession: T15969
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-619 <CH1>
A/Cross-references: EMBL:U23178; NID:G726421; PID:G726422; PIDN:AMC46719.1; CESP:F08B1.1
A/Experimental source: strain Bristol N2
C/Genetic: G
A/Gen: CESP:F08B1.1
A/Introns: 22/2; 57/1; 99/1; 160/3; 252/2; 549/3
Query Match 1.4%; Score 9; DB 2; Length 619;
Best local Similarity 100.0%; Pred. No. 1.9;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 243 HCLAGISRS 251
|||||||
Db 223 HCLAGISRS 231

RESULT 2
AB2353
hypothetical protein al14378 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AB2353
R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi, Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Ref. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AB2353
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1727 <KIR>
A:Cross-references: GB:BA000019; PIDN:BA076077.1; PID:gl7133514; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:

Query Match
Best Local Similarity 1.2%; Score 8; DB 2; Length 727;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 359 SVPSVPSV 366
|||||||
Db 116 SVPSVPSV 123

RESULT 3
T15280
hypothetical protein R155.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15280
R:Geisels, C.; Wamsley, P.; Kramet, J.
submitted to the EMBL Data Library, May 1997
A:Description: The sequence of C. elegans cosmid R155.
A:Reference number: Z18321
A:Accession: T15280
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1308 <GRI>
A:Cross-references: EMBL:AF003390; NID:g2088866; PID:g2088869; PIDN:AA054272.1; GSPDB:GN00179
A:Experimental source: strain Bristol N2; clone R155
C:Genetics:

A:Gene: CBSP:R155.2
A:Map position: 3
A:introns: 13/1; 57/1; 860/2; 897/1; 970/2; 1116/2; 1174/1; 1225/3

Query Match
Best Local Similarity 1.2%; Score 8; DB 2; Length 1308;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 247 GISRSATI 254
|||||||
Db 1165 GISRSATI 1172

RESULT 4
B70143
fibronectin/fibrinogen-binding protein homolog - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C:Accession: B70143

R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White son, D.; Peterson, J.; Keriavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt, ; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98065943; PMID:9403685
A:Accession: B70143
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-475 <KLE>
A:Cross-references: GB:AE001141; GB:AE000783; NID:g2688250; PIDN:AC66734.1; PID:g268825
A:Experimental source: strain B31

Query Match
Best Local Similarity 1.1%; Score 7; DB 2; Length 475;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 296 KKIRKQT 302
|||||||
Db 469 KKIRKQT 475

RESULT 5
S17836
fatty-acyl-CoA reductase (EC 1.2.1.-) luxC - Photobacterium leiognathi
C:Species: Photobacterium leiognathi
C:Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 26-May-2000
C:Accession: S17836
R:Lee, C.Y.; Sztitner, R.B.; Meighen, E.A.
Eur. J. Biochem. 201, 161-167, 1991
A:Title: The lux genes of the luminous bacterial symbiont, Photobacterium leiognathi, of coli.
A:Reference number: S17836; MUID:92007870; PMID:1915359
A:Accession: S17836
A:Molecule type: DNA
A:Residues: 1-478 <LEE>
A:Cross-references: EMBL:M63594; NID:g150687; PIDN:AA25616.1; PID:g150688
C:Genetics:
A:Gene: luxC
C:Superfamily: Photobacterium fatty-acyl-CoA reductase
C:Keywords: oxidoreductase

Query Match
Best Local Similarity 1.1%; Score 7; DB 2; Length 478;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 SVHLIAG 123
|||||||
Db 145 SVHLIAG 151

RESULT 6
T47415
transporter-like protein - Arabidopsis thaliana
N:Alternate names: protein T28A8 80
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T47415
R:Punelle, B.; Boutry, M.; Goffeau, A.; Mewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X
submitted to the Protein Sequence Database, March 2000
A:Reference number: Z24466
A:Accession: T47415
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-479 <PUR>
A:Cross-references: EMBL:AL162691
A:Experimental source: cultivar Columbia; BAC clone T28A8
C:Genetics:
A:Map position: 3
A:introns: 48/1; 77/1; 111/2; 145/3; 163/3; 184/3; 201/2; 226/3; 291/3; 321/1; 349/3; 38
A:Note: T28A8.80

```
Query Match      1.1%; Score 7; DB 2; Length 479;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      518 LFGGLSTS 524
      |||||
Db      117 LFGGLSTS 123

RESULT 7
AE0001
conserved hypothetical protein YPO0004 [imported] - Yersinia pestis (strain CO92)
C/Species: Yersinia pestis
C/Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #ext_change 09-Nov-2001
C/Accession: AE0001
R/Parhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
demo-Taruga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,
Nature 413, 523-527, 2001
A>Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A/Reference number: AB0001; MUID:21470413; PMID:11586360
A/Accession: AE0001
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-488 <KUR>
A/Cross-references: GB:AL50842; PIDN:CAC88871.1; PID:g15978119; GSPDB:GN00175
C/Genetics:
A/Genes: YPO0004
C/Superfamily: Escherichia coli hypothetical 49.6K protein (asna 3' region)

Query Match      1.1%; Score 7; DB 2; Length 488;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      369 SLEEDSP 375
      |||||
Db      80 SLEEDSP 86

RESULT 8
G84577
probable WD-40 repeat protein, MS14 [imported] - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #ext_change 02-Feb-2001
C/Accession: G84577
R/Idin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Xoo, H.; Moffatt, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A/Reference number: A84420; MUID:20083487; PMID:10617197
A/Accession: G84577
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-507 <STO>
A/Cross-references: GB:AE002093; NID:g4191782; PIDN:AAD10151.1; GSPDB:GN00139
C/Genetics:
A/Genes: A2g19520
A/Map position: 2

Query Match      1.1%; Score 7; DB 2; Length 507;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      302 TGASGPK 308
      |||||
Db      20 TGASGPK 26

RESULT 9
D86911
conserved hypothetical protein MU0020 [imported] - Mycobacterium leprae
C/Species: Mycobacterium leprae
C/Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #ext_change 20-Apr-2001
C/Accession: D86911
R/Cole, S.T.; Eigmeier, K.; Parhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; H.
R.; Davies, R.M.; Devlin, K.; Duthey, S.; Felwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.W.
Nature 409, 1007-1011, 2001
A/Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squires, R.; Sk
A>Title: Massive gene decay in the leprosy bacillus.
A/Reference number: A86909; MUID:21128732; PMID:11234002
A/Accession: D86911
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-509 <STO>
A/Cross-references: GB:AL450380; NID:g13092430; PIDN:CAC29528.1; GSPDB:GN00147
C/Genetics:
A/Genes: MU0020

Query Match      1.1%; Score 7; DB 2; Length 509;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      489 TSSSGTA 495
      |||||
Db      460 TSSSGTA 466

RESULT 10
T10013
probable phosphoprotein phosphatase (EC 3.1.3.16) - Mycobacterium leprae
C/Species: Mycobacterium leprae
C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #ext_change 16-Jul-1999
C/Accession: T10013
R/Cole, S.T.
Submitted to the EMBL Data Library, August 1997
A/Reference number: Z16916
A/Accession: T10013
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-509 <COL>
A/Cross-references: EMBL:Z70722; NID:e1059634; PID:e337960
C/Genetics:
A/Note: PPP
C/Keywords: phosphoric monoester hydrolase

Query Match      1.1%; Score 7; DB 2; Length 509;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      489 TSSSGTA 495
      |||||
Db      460 TSSSGTA 466

RESULT 11
A38337
amido-phosphoribosyltransferase (EC 2.4.2.14) precursor - chicken
N/Alternate names: glutamine PRP amido-transferase
C/Species: Gallus gallus (chicken)
C/Date: 28-Jun-1991 #sequence_revision 28-Jun-1991 #ext_change 18-Feb-2000
C/Accession: A38337; I50240
R/Zhou, G.; Dixon, J.E.; Zalkin, H.
J. Biol. Chem. 265, 21152-21159, 1990
A>Title: Cloning and expression of avian glutamine phosphoribosylpyrophosphate amidotrai
A/Reference number: A38337; MUID:91065928; PMID:2123487
A/Accession: A38337
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-510 <ZHO>
A/Cross-references: GB:M60069; GB:M8186; NID:g211824; PIDN:AAA62736.1; PID:g211825
R/Gavala, A.; Dixon, J.E.; Brayton, K.A.; Zalkin, H.
Mol. Cell. Biol. 13, 4784-4792, 1993
A>Title: Coexpression of two closely linked avian genes for purine nucleotide synthesis
A/Reference number: I50239; MUID:93330273; PMID:8336716
```

A:Accession: I50240
A:Status: preliminary; translated from GB/EMBL/DBD
A:Molecule type: DNA
A:Residues: 1-42 <GAV>
A:Cross-references: GB:LI2533; NID:G402688; PIDN:AAA17895.1; PID:G482248
C:Genetics:
A:Gene: GPAT
C:Superfamily: amidophosphoribosyltransferase
C:Keywords: 4fe-4S; glycosyltransferase; metalloprotein; pentosyltransferase
F:1-11/Domain: propeptide #status predicted <PRO>
F:12-510/Product: amidophosphoribosyltransferase #status predicted <MAT>
F:12/Active site: Cys #status predicted
F:280,426,496,499/Binding site: 4fe-4S cluster (Cys) (covalent) #status predicted

Query Match
Best Local Similarity 1.1%; Score 7; DB 2; Length 510;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 594 VVSRRR 600
| | | | |
| | | | |
DB 299 VVSRRR 305

RESULT 12
AE3405
hypothetical protein BME11227 [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: AE3405
R:Belicchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujar, C.; Loh, T.; Ivanova, P.; Mazur, M.; Goldsman, E.; Selkov, E.; Elker, P.H.; Hagius, S.; O'Callaghan, D.; Letessier, Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AD252; PMID:11756688
A:Accession: AE3405
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-511 <KUR>
A:Cross-references: GB:AE008917; PIDN:AL52408.1; PID:G17983210; GSPDB:GN00190
C:Genetics:
A:Experimental source: strain 16M
A:Gene: BME11227
A:Map position: 1

Query Match
Best Local Similarity 1.1%; Score 7; DB 2; Length 511;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 AKKVDI 79
| | | | |
| | | | |
DB 34 AKKVDI 40

RESULT 13
S61889
3-oxosteroid 1-dehydrogenase (EC 1.3.99.4) - Arthrobacter simplex
N:Alternate names: 3-ketosteroid-delta-dehydrogenase kedd
C:Species: Arthrobacter simplex
C>Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Jun-2000
C:Accession: S61889
R:Mohar, I.; Choi, K.P.; Yamashita, M.; Murooka, Y.
Mol. Microbiol. 15, 895-905, 1995
A:Title: Molecular cloning, expression in Streptomyces lividans, and analysis of a gene se and a hypothetical regulatory protein.
A:Reference number: S61887; MUID:9519331; PMID:7596291
A:Accession: S61889
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-515 <MOU>
A:Cross-references: EMBL:D37969; NID:G882006; PIDN:BA07106.1; PID:G1518167
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994
C:Genetics:
A:Note: only a part of the translation is shown

A:Gene: kedd
C:Superfamily: 3-oxosteroid 1-dehydrogenase; 3-oxosteroid 1-dehydrogenase homology
C:Keywords: oxidoreductase
F:403-501/Domain: 3-oxosteroid 1-dehydrogenase homology <OXD>

Query Match
Best Local Similarity 1.1%; Score 7; DB 2; Length 515;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 424 EDALRY 430
| | | | |
| | | | |
DB 71 EDALRY 77

RESULT 14
B97864
hypothetical protein hemk [imported] - Rickettsia conorii (strain Malish 7)
C:Species: Rickettsia conorii
C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 30-Sep-2001
C:Accession: B97864
R:Ogata, H.; Audic, S.; Renseto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; Ro Science 293, 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MUID:21442074; PMID:11557893
A:Accession: B97864
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-524 <KUR>
A:Cross-references: GB:AE006914; PIDN:AL03852.1; PID:G15620455; GSPDB:GN00173
C:Genetics:
A:Gene: hemk

Query Match
Best Local Similarity 1.1%; Score 7; DB 2; Length 524;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 178 ELIOQNG 184
| | | | |
| | | | |
DB 473 ELIOQNG 479

RESULT 15
S40461
t-complex-type molecular chaperone tcpl (clone ASTCP-K19) - oat
C:Species: Avena sativa (oat)
C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 20-Aug-1999
C:Accession: S40461; S39319
R:Elmann, B.; Krenz, M.; Mummert, E.; Schaefer, E.
FEBS Lett. 336, 313-316, 1993
A:Title: Two Tcpl-1-related but highly divergent gene families exist in oat encoding prot
A:Reference number: S40461; MUID:94085629; PMID:7903257
A:Accession: S40461
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-535 <EHM>
A:Cross-references: EMBL:X75777; NID:G435172; PIDN:CA53396.1; PID:G435173
A:Experimental source: cultivar pawi, tissue_type mesocotyl and coleoptile
C:Genetics:
A:Gene: tcpl
C:Superfamily: molecular chaperone t-complex-type
C:Keywords: molecular chaperone

Query Match
Best Local Similarity 1.1%; Score 7; DB 2; Length 535;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 KKKVDI 80
| | | | |
| | | | |
DB 259 KKKVDI 265

RESULT 16
S40462

t-complex-type molecular chaperone tcp1 (clone ASTCP-K36) - oat
C:Species: *Avena sativa* (oat)
C:Date: 19-May-1994 #sequence_revision 26-May-1995 #text_change.20-Aug-1999
C:Accession: S40462; S39320
R:Bhmann, B.; Krenz, M.; Mummert, E.; Schaefer, E.
FEBS Lett. 336, 313-316, 1993
A:Title: Two Tcp-1-related but highly divergent gene families exist in oat encoding pro
A:Reference number: S40461; MUID:94085629; PMID:7903257
A:Accession: S40462
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-535 <EMBL>
A:Cross-references: EMBL:X75778; NID:9435174; PIDN:CAA53397.1; PID:9435175
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1993
C:Genetics:
A:Gene: tcp1
C:Superfamily: molecular chaperone t-complex-type
C:Keywords: molecular chaperone

Query Match 1.1% Score 7; DB 2; Length 535;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 KHKVDID 80
DB 259 KHKVDID 265

RESULT 17
C72453
hypothetical protein APE2271 - Aeropyrum pernix (strain K1)
C:Species: *Aeropyrum pernix*
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jun-2000
C:Accession: C72453
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, *Aeropyr*
A:Reference number: A72450; MUID:99310339; PMID:10382966
A:Accession: C72453
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-546 <KAW>
A:Cross-references: DDBJ:AP00064; NID:95105945; PIDN:BA81283.1; PID:dl045069; PID:9510
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE2271
C:Superfamily: Aeropyrum pernix hypothetical protein APE2271

Query Match 1.1% Score 7; DB 2; Length 546;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 567 YSASAIY 573
DB 373 YSASAIY 379

RESULT 18
G81959
conserved hypothetical protein NMA0428 [imported] - *Neisseria meningitidis* (strain Z2491
C:Species: *Neisseria meningitidis*
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: G81959
R:Fairhill, J.; Achman, M.; James, K.D.; Bentley, S.D.; Churche, C.; Klee, S.R.; Morel
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491.
A:Reference number: A81775; MUID:20222556; PMID:10761919
A:Accession: G81959
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-548 <PAR>

A:Cross-references: GB:AL162753; GB:AL157959; NID:97379120; PIDN:CAB83727.1; PID:973791
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: NMA0428

Query Match 1.1% Score 7; DB 2; Length 548;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 562 ESSHFYS 568
DB 281 ESSHFYS 287

RESULT 19
F72268
ABC transporter, ATP-binding protein - *Thermotoga maritima* (strain MSB8)
C:Species: *Thermotoga maritima*
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: F72268
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey
Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome seq
A:Reference number: A72200; MUID:99287316; PMID:10360571
A:Accession: F72268
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-564 <ARN>
A:Cross-references: GB:AE001786; GB:AE000512; NID:94981873; PIDN:AAD36392.1; PID:949818
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM1319
C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology

Query Match 1.1% Score 7; DB 2; Length 564;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 137 ESKSTLY 143
DB 367 ESKSTLY 373

RESULT 20
S50979
RHC21 protein - yeast (*Saccharomyces cerevisiae*)
N:Alternate names: protein D2940; protein YD8119_04; protein YDL003w
C:Species: *Saccharomyces cerevisiae*
C:Date: 11-Feb-1995 #sequence_revision 12-May-1995 #text_change 05-Dec-1997
C:Accession: S50979; S5719; S52516; S67535
R:Murphy, L.; Richards, C.; Gentile, S.; Harris, D.
submitted to the EMBL Data Library, January 1995
A:Reference number: S50976
A:Accession: S50979
A:Molecule type: DNA
A:Residues: 1-566 <MUR>
A:Cross-references: EMBL:Z48008; NID:9642799; PID:9642803
R:Strunnikov, A.V.; Koshland, D.
submitted to the EMBL Data Library, March 1995
A:Reference number: S55719
A:Accession: S55719
A:Molecule type: DNA
A:Residues: 1-566 <STR>
A:Cross-references: EMBL:U23759; NID:9777397; PID:9777398
R:Andre, B.; Vissers, S.; Ureastarazu, L.
submitted to the EMBL Data Library, February 1995
A:Description: The sequence of a 42 kb segment located on the left arm of chromosome IV
A:Reference number: S52492
A:Accession: S52516
A:Molecule type: DNA
A:Residues: 1-285 <AND>

A/Cross-references: EMBL:Z48432; NID:9683669; PID:9683694
 A/Experimental source: strain S288C
 R/Ureastarazu, L.A.; Andre, B.; Vissers, S.
 submitted to the Protein Sequence Database, July 1996
 A/Reference number: 567535
 A/Accession: 567535
 A/Molecule type: DNA
 A/Residues: 1-285 <URR>
 A/Cross-references: EMBL:Z74051; MIPS:YDL003w
 A/Experimental source: strain S288C
 C/Genetics:
 A/Gene: SGD:MCD1; RHC21
 A/Cross-references: MIPS:YDL003w; SGD:S0002161
 A/Map position: 4L

Query Match 1.1% Score 7; DB 2; Length 566;
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 633 GESISMS 639
 Db 271 GESISMS 277

RESULT 21
 JTO494
 alpha-glucosidase homolog precursor - yellow fever mosquito
 N/Alternate names: maltase homolog
 C/Species: Aedes aegypti (yellow fever mosquito)
 C/Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 22-Jun-1999
 C/Accession: JTO494
 R/James, A.A.; Blackmer, K.; Racioppi, J.V.
 Gene 75, 73-83, 1989
 A/Title: A salivary gland-specific, maltase-like gene of the vector mosquito, Aedes aegypti
 A/Reference number: JTO494; MID:89252923; PMID:2470653
 A/Accession: JTO494
 A/Molecule type: mRNA
 A/Residues: 1-579 <UAM>
 A/Cross-references: GB:M23222; GB:M23264; NID:9159565; PIDN:AAA29352.1; PID:9
 A/Note: the authors translated the codon CAA for residue 569 as Lys
 C/Genetics:
 A/Gene: Mali
 C/Superfamily: alpha-glucosidase; alpha-amylase core homology
 C/Keywords: glycoprotein
 F/1-18/Domain: signal sequence #status predicted <SIG>
 F/19-579/Product: alpha-glucosidase homolog #status predicted <MAT>
 F/187-359/Domain: alpha-amylase core homology <AMY>
 F/118,151,282,304,325,401/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 1.1% Score 7; DB 2; Length 579;
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 577 ASYSAYS 583
 Db 414 ASYSAYS 420

RESULT 22
 T11624
 spindle poison sensitivity protein - fission yeast (Schizosaccharomyces pombe)
 C/Species: Schizosaccharomyces pombe
 C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
 C/Accession: T11624
 R/Bartell, B.G.; Rajandream, M.A.; Wood, V.
 submitted to the EMBL Data Library, September 1997
 A/Reference number: Z17303
 A/Accession: T11624
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-583 <BAR>
 A/Cross-references: EMBL:Z99260; NID:e1061997; PID:e1341400
 C/Genetics:

A/Map position: IL
 A/Introns: 35/1
 A/Note: SPAC3A11.02

Query Match 1.1% Score 7; DB 2; Length 583;
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 408 SYSASMA 414
 Db 238 SYSASMA 304

RESULT 23
 S55205
 dihydroxy-acid dehydratase (EC 4.2.1.9) - yeast (Saccharomyces cerevisiae)
 N/Alternate names: protein J1450; protein YJR016c
 C/Species: Saccharomyces cerevisiae
 C/Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 22-Jun-1999
 C/Accession: S55205; S57031; S43744
 R/de Haan, M.; Smits, P.H.M.; Grivell, L.A.
 submitted to the EMBL Data Library, May 1995
 A/Reference number: S55183
 A/Accession: S55205
 A/Molecule type: DNA
 A/Residues: 1-585 <DEH>
 A/Cross-references: EMBL:X87611; NID:9854567; PIDN:CAA60939.1; PID:9854590
 R/de Haan, M.; Grivell, L.A.; Smits, P.H.M.
 submitted to the Protein Sequence Database, September 1995
 A/Reference number: S56771
 A/Accession: S57031
 A/Molecule type: DNA
 A/Residues: 1-585 <ZAG>
 A/Cross-references: EMBL:Z49516; NID:91015649; PIDN:CAA89540.1; PID:91015650; MIPS:YJR01
 R/Velasco, J.A.; Canedo, J.; Pena, M.C.; Kawakami, T.; Laborda, J.; Notario, V.
 submitted to the EMBL Data Library, January 1994
 A/Description: Cloning of the dihydroxyacid dehydratase gene (ILV-3) from the yeast Sacc
 A/Reference number: S43744
 A/Accession: S43744
 A/Molecule type: DNA
 A/Residues: 'MF', 84-85, 'IEKR', 90-237, 'S', 239-241, 'S', 243-320, 'P', 322-452 <VEL>
 A/Cross-references: EMBL:L13975; NID:9439111; PIDN:AAA34568.1; PID:9553125
 C/Genetics:
 A/Gene: SGD:ILV3
 A/Cross-references: SGD:S0003777; MIPS:YJR016c
 A/Map position: 10R
 C/Superfamily: dihydroxy-acid dehydratase
 C/Keywords: carbon-oxygen lyase; hydro-lyase

Query Match 1.1% Score 7; DB 2; Length 585;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 235 ASNGCVL 241
 Db 577 ASNGCVL 583

RESULT 24
 T17333
 hypothetical protein DKFZp434E165.1 - human (fragment)
 C/Species: Homo sapiens (man)
 C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C/Accession: T17333
 R/Diescherhoef, A.; Lauber, J.; Neues, H.W.; Gassenhuber, J.; Wiemann, S.
 submitted to the Protein Sequence Database, September 1999
 A/Reference number: Z18727
 A/Accession: T17333
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-596 <DUE>
 A/Cross-references: EMBL:AL117632
 A/Experimental source: adult testis; clone DKFZp434E165

C:Genetics:
A:Note: DKFP434E165.1

Query Match
Best Local Similarity 1.1%; Score 7; DB 2; Length 596;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 549 STPLSTS 555
|||||
Db 388 STPLSTS 394

RESULT 25
T35054
probable transport system permease protein - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C:Accession: T35054
R:Seeger, K.J.; Harris, D.; Parkhill, J.; Barrrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, August 1998
A:Reference number: Z21566
A:Accession: T35054
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-601 <SEB>
A:Cross-references: EMBL:AL031371; PIDN:CAA20556.1; GSPDB:GN00070; SCOEDB:SC4G2.19
C:Genetics:
A:Gene: SCOEDB:SC4G2.19

Query Match
Best Local Similarity 1.1%; Score 7; DB 2; Length 601;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 349 AACGRPV 355
|||||
Db 582 AACGRPV 588

RESULT 26
G90575
transketolase (tk) [imported] - Mycoplasma pulmonis (strain UAB CRIP)
C:Species: Mycoplasma pulmonis
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C:Accession: G90575
R:Chandaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A:Reference number: A93512; MUID:21267165; PMID:11353084
A:Accession: G90575
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-615 <KUR>
A:Cross-references: GB:AL445566; PID:G1408925; PIDN:CAC13684.1; GSPDB:GN00153
A:Experimental source: strain UAB CRIP
C:Genetics:
A:Gene: MYPV 5110
A:Genetic code: SGC3
C:Superfamily: transketolase; thiamin pyrophosphate-binding domain homology

Query Match
Best Local Similarity 1.1%; Score 7; DB 2; Length 615;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 230 IEKAKAS 236
|||||
Db 222 IEKAKAS 228

RESULT 27
B86431
hypothetical protein TS18.4 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
C:Accession: B86431
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso
Chin, C.W.; Chung, W.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.
ansen, N.F.; Hughes, B.; Hutzler, L.
Nature 408, 916-920, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luroe, J.S.; Malt, R.; Marziani
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: B86431
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-622 <STO>
A:Cross-references: GB:AE005172; NID:G4587515; PIDN:AAD25746.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match
Best Local Similarity 1.1%; Score 7; DB 2; Length 622;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 281 ISPNFNF 287
|||||
Db 537 ISPNFNF 543

RESULT 28
A39262
transcription factor HNF-1A - mouse
N:Alternate names: hepatic nuclear factor 1-alpha; transcription factor APF; transcript:
C:Species: Mus musculus (house mouse)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A39262; S70436
R:Kuo, C.T.; Conley, P.B.; Hsieh, C.L.; Francke, U.; Crabtree, G.R.
Proc. Natl. Acad. Sci. U.S.A. 87, 9838-9842, 1990
A:Title: Molecular cloning, functional expression, and chromosomal localization of mouse
A:Reference number: A39262; MUID:91088607; PMID:2263635
A:Accession: A39262
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-628 <Kuo>
A:Cross-references: GB:M57966; NID:G193885; PIDN:AAA37821.1; PID:G193886
R:Bach, I.; Pontoglio, M.; Yaniv, M.
Nucleic Acids Res. 20, 4199-4204, 1992
A:Title: Structure of the gene encoding hepatocyte nuclear factor 1 (HNF1).
A:Reference number: S70435; MUID:92375726; PMID:1354855
A:Accession: S70436
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 199-279 <BAC>
A:Note: only a part of the nucleic acid sequence is shown
C:Genetics:
A:Gene: Hnf-1
A:Introns: 238/2
A:Note: the list of introns is incomplete
C:Complex: homodimer; can also form heterodimers with, for example, HNF-1B
C:Function:
A:Description: transcription activator required for the expression of a number of liver.
A:Note: also expressed in other tissues (kidney, stomach, intestine), where it may play
C:Superfamily: transcription factor HNF-1; homeobox homology
C:Keywords: DNA binding; heterodimer; homeobox; homodimer; liver; nucleus; transcription
F:1-33/Region: dimerization
F:222-277/Domain: homeobox homology <Hox>

Query Match
Best Local Similarity 1.1%; Score 7; DB 1; Length 628;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 344 SATSENA 350

Db 325 SATSEAA 331

RESULT 29

transcription factor HNF-1A - rat
 N:Alternate names: hepatic nuclear factor 1-alpha; transcription factor APF; transcript
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: A33333; A34590; S11568; S70435
 R:Pratt, M.; Swart, G.; Monaci, P.; Nicotri, A.; Staemphi, S.; Frank, R.; Cortese, R.
 Cell 59, 145-157, 1989
 A>Title: The liver-specific transcription factor LF-B1 contains a highly diverged homeo
 A:Reference number: A33333; UID:90003224; PMID:2571419
 A:Accession: A33333
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-628 <PRA>
 A:Cross-references: GB:J03170; NID:9205164; PIDN:AAA1524.1; PID:9205165
 R:Baumhueter, S.; Mendel, D.B.; Conley, P.B.; Kuo, C.J.; Turk, C.; Graves, M.K.; Edwards
 Genes Dev. 4, 372-379, 1990
 A>Title: HNF-1 shares three sequence motifs with the POU domain proteins and is identica
 A:Reference number: A34590; UID:90249741; PMID:1970973
 A:Accession: A34590
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 'R', 166-628 <BAU>
 A:Cross-references: GB:X53297; NID:957867; PIDN:CAA37387.1; PID:957868
 A:Note: the authors translated the codon GAG for residue 616 as Asp
 R:Chouard, T.; Blumenfeld, M.; Bach, I.; Vandekerckhove, J.; Cereghini, S.; Yaniv, M.
 Nucleic Acids Res. 18, 5853-5863, 1990
 A>Title: A distal dimerization domain is essential for DNA-binding by the atypical HNF1
 A:Reference number: S11568; UID:91016926; PMID:2216777
 A:Accession: S11568
 A:Molecule type: mRNA
 A:Residues: 1-628 <CHO>
 A:Cross-references: EMBL:X54423; NID:956367; PIDN:CAA38295.1; PID:956368
 R:Bach, I.; Fontoglio, M.; Yaniv, M.
 Nucleic Acids Res. 20, 4199-4204, 1992
 A>Title: Structure of the gene encoding hepatocyte nuclear factor 1 (HNF1).
 A:Reference number: S70435; UID:92375726; PMID:1354855
 A:Accession: S70435
 A:Status: not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 199-279 <BAC>
 A:Note: only a part of the nucleic acid sequence is shown
 C:Genetics:
 A:Gene: Hnf-1
 A:Introns: 238/2
 A:Note: the list of introns is incomplete
 C:Complex: homodimer; can also form heterodimers with, for example, HNF-1B
 C:Function:
 A:Description: transcription activator required for the expression of a number of liver-
 A:Note: also expressed in other tissues (kidney, stomach, intestine), where it may play
 C:Superfamily: transcription factor HNF-1; homeobox homology
 C:Keywords: DNA binding; heterodimer; homeobox; homodimer; liver; nucleus; transcription
 F:1-33/Region: dimerization
 F:222-277/Domain: homeobox homology <HOX>

Query Match 1.1%; Score 7; DB 1; Length 628;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 344 SATSEAA 350
 Db 325 SATSEAA 331

RESULT 30
 G90779
 hypothetical protein Eca1207 [imported] - Escherichia coli (strain O157:H7, substrain RI
 C:Species: Escherichia coli

C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
 C:Accession: G90779
 R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 Gasawara, N.; Yasunaga, T.; Kihara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
 A:Reference number: A9629; UID:21156231; PMID:11258796
 A:Accession: G90779
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-634 <HAV>
 A:Cross-references: GB:BA000007; PIDN:BA834630.1; PID:913360667; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain RIMD 0509952
 C:Genetics:
 A:Gene: Eca1207

Query Match 1.1%; Score 7; DB 2; Length 634;
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 417 LHGESS 423
 Db 617 LHGESS 623

RESULT 31

hypothetical protein - human herpesvirus 4
 C:Species: human herpesvirus 4, Epstein-Barr virus
 C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 13-Jan-1995
 C:Accession: S32991
 R:Farrell, P.J.
 submitted to the EMBL Data Library, March 1988
 A:Reference number: S32973
 A:Accession: S32991
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-657 <FAR>
 A:Cross-references: EMBL:V01555

Query Match 1.1%; Score 7; DB 2; Length 657;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 324 PAVSEGG 330
 Db 379 PAVSEGG 385

RESULT 32

F81896
 probable phage transposase NMA1284 [imported] - Neisseria meningitidis (strain Z2491 ser
 C:Species: Neisseria meningitidis
 C>Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
 C:Accession: F81896
 R:Farhah, J.; Achtmann, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
 ; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,
 Nature 404, 502-506, 2000
 A>Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.
 A:Reference number: A81775; UID:20222556; PMID:10761919
 A:Accession: F81896
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-657 <PAR>
 A:Cross-references: GB:AL162755; GB:AL157959; NID:97379742; PIDN:CA884536.1; PID:9737996
 A:Experimental source: serogroup A, strain Z2491
 C:Genetics:
 A:Gene: NMA1284

Query Match 1.1%; Score 7; DB 2; Length 657;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 TERLVAL 18.
|||||
Db 178 TERLVAL 184

RESULT 33
A81125
bacteriophage transposase NMB1081 [imported] - Neisseria meningitidis (strain MCS8 serog
C/Species: Neisseria meningitidis
C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C/Accession: A81125
R/RetelIn, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A
Hickey, E.K.; Haft, D.H.; Salberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.
A/Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A/Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MCS8.
A/Reference number: A81000; MUID:20175755; PMID:10710307
A/Accession: A81125
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-661 <TET>
A/Cross-references: GB:AE002458; GB:AE002098; NID:g7226311; PIDN:AAF41473.1; PID:g722631
A/Experimental source: serogroup B, strain MCS8
C/Genetics:
A/Gene: NMB1081

Query Match 1.1%; Score 7; DB 2; Length 661;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 12 TERLVAL 18
|||||
Db 182 TERLVAL 188

RESULT 34
S45131
probable membrane protein YNR008w - yeast (Saccharomyces cerevisiae)
N/Alternate names: hypothetical protein N2042
C/Species: Saccharomyces cerevisiae
C/Date: 03-May-1994 #sequence_revision 02-Aug-1994 #text_change 19-Apr-2002
R/Verhasselt, P.; Aert, R.; Voet, M.; Volckaert, G.
A/Accession: S45131; S48345; S63334
Submitted to the EMBL Data Library, January 1994
A/Description: Twelve open reading frames revealed on the 23.6 kbp segment flanking the
A/Reference number: S45119
A/Accession: S45131
A/Molecule type: DNA
A/Residues: 1-661 <YBR>
A/Cross-references: EMBL:X77395; NID:g496717; PID:g496725
R/Verhasselt, P.; Aert, R.; Voet, M.; Volckaert, G.
Yeast 10, 1355-1361, 1994
A/Title: Twelve open reading frames revealed in the 23.6 kb segment flanking the centrom
A/Reference number: S48338; MUID:95208356; PMID:7900425
A/Accession: S48345
A/Status: nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-661 <YB2>
A/Cross-references: EMBL:X77395; NID:g496717; PIDN:CA54576.1; PID:g496725
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1994
R/Aert, R.; Verhasselt, P.; Voet, M.; Volckaert, G.
Submitted to the Protein Sequence Database, April 1996
A/Reference number: S62910
A/Accession: S63334
A/Molecule type: DNA
A/Residues: 1-661 <AER>
A/Cross-references: EMBL:X71623; NID:g1302481; PID:e239782; PID:g1302482; MIPS:YNR008w
A/Experimental source: strain S288C
C/Genetics:
A/Gene: SGD:LR01
A/Cross-references: SGD:S0005291
A/Map position: 14R

C/Keywords: transmembrane protein
F:80-96/Domain: transmembrane #status predicted <TM>

Query Match 1.1%; Score 7; DB 2; Length 661;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 421 SSSEPAL 427
|||||
Db 431 SSSEPAL 437

RESULT 35
C81505
hypothetical protein CP1075 [imported] - Chlamydia pneumoniae (strain AR39)
C/Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 11-May-2000
C/Accession: C81505
R/Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwin, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salberg
Nucleic Acids Res. 28, 1397-1406, 2000
A/Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A/Reference number: A81500; MUID:20150255; PMID:10684935
A/Accession: C81505
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-674 <REA>
A/Cross-references: GB:AE002264; GB:AE002161; NID:g7189984; PIDN:AAF38847.1; PID:g718998
A/Experimental source: strain AR39, HL cells
C/Genetics:
A/Gene: CP1075

Query Match 1.1%; Score 7; DB 2; Length 674;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 LIQHSAX 74
|||||
Db 414 LIQHSAX 420

RESULT 36
E72033
hypothetical protein - Chlamydia pneumoniae (strain CWL029)
C/Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C/Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000
C/Accession: E72033
R/Kalman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A/Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A/Reference number: A72000; MUID:9206606; PMID:10192388
A/Accession: E72033
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-680 <ARN>
A/Cross-references: GB:AE001661; GB:AE001363; NID:g4377104; PIDN:AAD18934.1; PID:g437710
A/Experimental source: strain CWL029
C/Genetics:
A/Gene: CPn0796

Query Match 1.1%; Score 7; DB 2; Length 680;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 LIQHSAX 74
|||||
Db 420 LIQHSAX 426

RESULT 37
B86590
hypothetical protein CPJ0796 [imported] - Chlamydia pneumoniae (strain J138)
C/Species: Chlamydia pneumoniae, Chlamydia pneumoniae

C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
 C/Accession: B86590
 R/Shiiri, M.; Hirakawa, H.; Kinoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Is
 Nucleic Acids Res. 29, 2311-2314, 2000
 A>Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
 A/Reference number: A86491; MUID:20330349; PMID:10871362
 A/Accession: B86590
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-680 <STO>
 A/Cross-references: GB:BA000008; NID:g9879169; PIDN:BAA99004.1; GSPDB:GN00142
 A/Experimental source: strain J138
 C/Genetics:
 A/Gene: Cpj0796

Query Match 1.1%; Score 7; DB 2; Length 680;
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 68 LIQHSK 74
 |||||
 Db 420 LIQHSK 426

RESULT 38
 G82618
 plus biogenesis protein Xp1953 [imported] - Xylella fastidiosa (strain 9asc)
 C/Species: Xylella fastidiosa
 C/Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
 C/Accession: G82618
 R/anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
 Nature 406, 151-157, 2000
 A>Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A/Reference number: A82515; MUID:20355717; PMID:10910347
 A/Note: for a complete list of authors see reference number A59328 below
 A/Accession: G82618
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-693 <SIM>
 A/Cross-references: GB:AE004014; GB:AE003849; NID:g9107044; PIDN:AAF64755.1; GSPDB:GN001

A/Experimental source: strain 9asc
 R/Simpson, A.J.G.; Reinach, F.C.; Arrudá, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
 Brito, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carrau, D.M.; Carreir, H
 A/Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
 submitted to Genbank, June 2000
 A/Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laiz
 Chado, M.A.; Madella, A.M.B.N.; Madella, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, B
 A/Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmiter, D.A
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
 A/Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
 M.; Tuhato, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
 A/Reference number: A59328
 A/Contents: annotation
 C/Genetics:
 A/Gene: Xf1953

Query Match 1.1%; Score 7; DB 2; Length 693;
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 390 LEDSNKL 396
 |||||
 Db 295 LEDSNKL 301

RESULT 39
 S01991
 transforming protein B-myb - human
 C/Species: Homo sapiens (man)
 C/Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 18-Jun-1999
 C/Accession: S01991; S48661

R/Nomura, N.; Takahashi, M.; Matsui, M.; Ishii, S.; Date, T.; Saeemoto, S.; Ishizaki, R
 Nucleic Acids Res. 16, 11075-11089, 1988
 A>Title: Isolation of human cDNA clones of myb-related genes, A-myb and B-myb.
 A/Reference number: S01991; MUID:89083548; PMID:3060855
 A/Accession: S01991
 A/Molecule type: mRNA
 A/Residues: 1-700 <NON>
 A/Cross-references: EMBL:X13293; NID:g29471; PIDN:CAA31655.1; PID:g29472
 R/Takemoto, Y.; Tashiro, S.; Handa, H.; Ishii, S.
 FEBS Lett. 350, 55-60, 1994
 A>Title: Multiple nuclear localization signals of the B-myb gene product.
 A/Reference number: S48661; MUID:94341374; PMID:8062924
 A/Contents: annotation; identification of nuclear location signal regions by mutational
 C/Genetics:
 A/Gene: GDB:MYB12; BMYB
 A/Cross-references: GDB:128109; OMIM:310305; OMIM:601415
 A/Map position: Xq13-Xq13
 C/Function:
 A/Description: transcription regulation; widespread activator of cell cycle genes; repre
 C/Superfamily: myb transforming protein; myb DNA-binding repeat homology
 C/Keywords: DNA binding; duplication; nucleus; transcription regulation
 F:26-77/Domain: myb DNA-binding repeat homology <MYB1>
 F:78-129/Domain: myb DNA-binding repeat homology <MYB2>
 F:130-180/Domain: myb DNA-binding repeat homology <MYB3>
 F:411-417/Region: nuclear location signal
 F:569-584/Region: nuclear location signal

Query Match 1.1%; Score 7; DB 1; Length 700;
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 CSQKVV 87
 |||||
 Db 479 CSQKVV 485

RESULT 40
 S33704
 transforming protein B-myb - mouse
 C/Species: Mus musculus (house mouse)
 C/Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 18-Jun-1999
 C/Accession: S33704; I48269
 R/Lam, E.W.F.; Robinson, C.; Watson, R.J.
 Oncogene 7, 1885-1890, 1992
 A>Title: Characterization and cell cycle-regulated expression of mouse B-myb.
 A/Reference number: S33704; MUID:92366176; PMID:1501895
 A/Accession: S33704
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-704 <LAM>
 A/Cross-references: EMBL:X70472; NID:g312825; PIDN:CAA49898.1; PID:g312826
 R/Lam, E.W.; Watson, R.J.
 EMBO J. 12, 2705-2713, 1993
 A>Title: An E2F-binding site mediates cell-cycle regulated repression of mouse B-myb tra
 A/Reference number: I48269; MUID:93327760; PMID:8334969
 A/Accession: I48269
 A/Status: preliminary; translated from GB/EMBL/DD83
 A/Molecule type: DNA
 A/Residues: 1-6 <RES>
 A/Cross-references: EMBL:X73028; NID:g297153; PIDN:CAA51511.1; PID:g581921
 C/Genetics:
 A/Gene: B-myb
 C/Function:
 A/Description: transcription regulation; widespread activator of cell cycle genes; repre
 C/Superfamily: myb transforming protein; myb DNA-binding repeat homology
 C/Keywords: DNA binding; duplication; nucleus; transcription regulation
 F:26-77/Domain: myb DNA-binding repeat homology <MYB1>
 F:78-129/Domain: myb DNA-binding repeat homology <MYB2>
 F:130-180/Domain: myb DNA-binding repeat homology <MYB3>
 F:414-420/Region: nuclear location signal
 F:573-588/Region: nuclear location signal

Query Match 1.1%; Score 7; DB 1; Length 704;

Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 CSQKVV 87
DB 482 CSQKVV 488

RESULT 41

T38117
probable protein kinase - fission yeast (Schizosaccharomyces pombe)

C/Species: Schizosaccharomyces pombe

C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C/Accession: T38117

R/Comment: R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.

submitted to the EMBL Data Library, August 1996

A/Reference number: Z21771

A/Accession: T38117

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1704 <CON>

A/Cross-references: EMBL:Z298600; PIDN:CAB1253.1; GSPDB:GN00066; SPDB:SPAC20G4.03C

A/Experimental source: strain 972h-; cosmid c20G4

C/Genetic: A/Map position: 1

Query Match 1.1%; Score 7; DB 2; Length 704;

Best Local Similarity 100.0%; Pred. No. 2.5e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 LVALLS 21
DB 50 LVALLS 56

RESULT 42

T17002
probable beta-galactosidase (EC 3.2.1.23) precursor - apple tree

C/Species: Malus domestica (apple tree)

C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000

C/Accession: T17002

R/Ross, G.S.; Megryzn, T.; MacRae, E.A.; Redgwell, R.J.

Plant Physiol. 106, 521-528, 1994

A/Title: Apple beta-galactosidase. Activity against cell wall polysaccharides and charac

A/Reference number: Z18645; MUID:95083752; PMID:7991682

A/Accession: T17002

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1731 <ROS>

A/Cross-references: EMBL:L29451; NID:g507277; PIDN:AAA62324.1; PID:g507278

A/Experimental source: strain Granny Smith; cortical

C/Superfamily: beta-galactosidase bga

C/Keywords: glycosidase; hydrolase

F1-23/Domain: signal sequence #status predicted <SIG>

F1-74-731/Product: probable beta-galactosidase #status predicted <MAT>

Query Match 1.1%; Score 7; DB 2; Length 731;

Best Local Similarity 100.0%; Pred. No. 2.6e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 648 KVGSSS 654
DB 414 KVGSSS 420

RESULT 43

C99995
hypothetical protein Ec82931 [imported] - Escherichia coli (strain O157:H7, substrain RI

C/Species: Escherichia coli

C/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 17-May-2002

C/Accession: C99995

R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen

A/Reference number: A99629; MUID:21156231; PMID:11238796

A/Accession: C99995

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1745 <HAY>

A/Cross-references: GB:BA000007; PIDN:BA836354.1; PID:g13362400; GSPDB:GN00154

A/Experimental source: strain O157:H7, substrain RIMD 0509952

C/Genetic: A/Supfamily: Escherichia coli hypothetical 68.5K protein (molR-DglX region)

Query Match 1.1%; Score 7; DB 2; Length 745;

Best Local Similarity 100.0%; Pred. No. 2.7e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 417 LHGFSS 423
DB 728 LHGFSS 734

RESULT 44

H85840

hypothetical protein yehQ [imported] - Escherichia coli (strain O157:H7, substrain EDL93

C/Species: Escherichia coli

C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 03-Jun-2002

C/Accession: H85840

R/Ferna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; Mayner

iller, L.; Grobeck, E.J.; Davis, N.W.; Llm, A.; Dimalanta, E.; Potamouats, K.; Apodaca,

Nature 409, 529-533, 2001

A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A/Reference number: A85480; MUID:21074935; PMID:11206551

A/Accession: H85840

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1745 <STO>

A/Cross-references: GB:AE005174; NID:g12516346; PIDN:AA657188.1; GSPDB:GN00145; UNGP:23

A/Experimental source: strain O157:H7, substrain EDL933

C/Genetic: A/Supfamily: yehQ

C/Superfamily: Escherichia coli hypothetical 68.5K protein (molR-DglX region)

Query Match 1.1%; Score 7; DB 2; Length 745;

Best Local Similarity 100.0%; Pred. No. 2.7e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 417 LHGFSS 423
DB 728 LHGFSS 734

RESULT 45

E82013

probable transcription accessory protein MMA0194 [imported] - Neisseria meningitidis (ser

C/Species: Neisseria meningitidis

C/Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001

C/Accession: E82013

R/Farhidi, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel

; Holroyd, S.; Jagels, K.; Leather, S.; Mowle, S.; Mungall, K.; Quail, M.A.; Rajandream,

Nature 404, 502-506, 2000

A/Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.

A/Reference number: AB1775; MUID:20222556; PMID:10761919

A/Accession: E82013

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1757 <PAR>

A/Cross-references: GB:AL162752; GB:AL157959; NID:g7378778; PIDN:CAB83508.1; PID:g737896

A/Experimental source: serogroup A, strain Z2491

C/Genetic: A/Supfamily: hypothetical protein ydcI

C:Keywords: methyltransferase; S-adenosylmethionine

Query Match 1.1%; Score 7; DB 2; Length 757;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 638 SENSRSR 644
|||||
DB 723 SENSRSR 729

RESULT 46

A81242

transcription accessory protein Tex, probable NMB0075 [imported] - Neisseria meningitidis

C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001

C:Accession: A81242

R:Rettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
Hickey, E.K.; Haft, D.H.; Salberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
H. Jin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.
Science 287, 1809-1815, 2000

A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:2015755; PMID:10710307

A:Accession: A81242

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-757 <TEXT>

A:Cross-references: GB:AE002366; GB:AE002098; NID:g7225284; PIDN:AAF40542.1; PID:g722529

A:Experimental source: serogroup B, strain MC58

C:Genetics:

A:Gene: NMB0075

C:Superfamily: hypothetical protein ydc1

Query Match 1.1%; Score 7; DB 2; Length 757;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 638 SENSRSR 644
|||||
DB 723 SENSRSR 729

RESULT 47

E97224

ATP-dependent Lon protease [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 02-Jun-2003

C:Accession: E97224

R:Nolling, U.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: E97224

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-778 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK80584.1; PID:g15025664; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC2637

C:Superfamily: ATP-dependent Lon protease

Query Match 1.1%; Score 7; DB 2; Length 778;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 659 MEIIEVS 665
|||||
DB 484 MEIIEVS 490

RESULT 48

T51759

[glutamate-ammonia-lyase] adenylyltransferase (EC 2.7.7.42) [imported] - Streptomyces c

C:Species: Streptomyces coelicolor

C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 03-Jun-2002

C:Accession: T51759

R:Fink, D.; Falke, D.; Mohlleben, W.; Engels, A.
Microbiology 145, 2313-2322, 1999

A:Title: Nitrogen metabolism in Streptomyces coelicolor A3(2): modification of glutamine

A:Reference number: 225448

A:Accession: T51759

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-784 <FIN>

A:Cross-references: EMBL:Y17736; PIDN:CAI6840.2

C:Genetics:

A:Gene: glnB

C:Keywords: nucleotidyltransferase

Query Match 1.1%; Score 7; DB 2; Length 784;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 267 IDEAYRF 273
|||||
DB 410 IDEAYRF 416

RESULT 49

T34468

hypothetical protein ZK770.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T34468

R:Maggi, L.; Gattung, S.; Bartko, L.
submitted to the EMBL Data Library, April 1997

A:Description: The sequence of C. elegans cosmid ZK770.

A:Reference number: 221530

A:Accession: T34468

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-795 <MAG>

A:Cross-references: EMBL:U97404; PIDN:AA93309.1; GSPDB:GN00019; CESP:ZK770.1

A:Experimental source: strain Bristol N2; clone ZK770

C:Genetics:

A:Gene: CESP:ZK770.1

A:Map position: 1

A:introns: 18/1; 84/2; 125/3; 198/3; 257/3; 337/3; 378/3; 424/3; 477/2; 510/1; 563/3; 63

Query Match 1.1%; Score 7; DB 2; Length 795;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 406 SVSYSAS 412
|||||
DB 665 SVSYSAS 671

RESULT 50

T03544

hypothetical protein - Rhodobacter capsulatus

C:Species: Rhodobacter capsulatus

C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 08-Oct-1999

C:Accession: T03544

R:Vlcek, C.; Paces, V.; Maltsev, N.; Paces, J.; Haselkorn, R.; Fonstein, M.
Proc. Natl. Acad. Sci. U.S.A. 94, 9384-9388, 1997

A:Title: Sequence of a 189-bp segment of the chromosome of Rhodobacter capsulatus SB1003

A:Reference number: Z14955; MUID:97404404; PMID:9256491

A:Accession: T03544

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-828 <VLC>

A:Cross-references: EMBL:AF010496; NID:g3128256; PIDN:AAI6197.1; PID:g3128345

C:Genetics:
A:Map position: 1

Query Match 1.1%; Score 7; DB 2; Length 828;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 119 HLAGGF 125
|||
Db 666 HLAGGF 672

Search completed: June 21, 2004, 13:28:58
Job time : 25 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 13:22:43 ; Search time 17 Seconds

(without alignments)
2036.862 Million cell updates/sec

Title: US-10-029-345A-109

Perfect score: 665
Sequence: 1 MAHEMIGTIVTERLVALL.....LGVSGSSFSGSMETIEVS 665

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 33406

Minimum DB seq length: 473

Maximum DB seq length: 200000000

Post-processing: Listing first 75 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	472	71.0	665	1 DUSG_HUMAN	Q9bY84 homo sapien
2	20	3.0	625	1 DUS8_HUMAN	Q13202 homo sapien
3	20	3.0	663	1 DUS8_MOUSE	Q09112 mus musculu
4	9	1.4	619	1 VHP1_CAEEL	Q10038 caenorhabdi
5	7	1.1	473	1 ATPB_PTEBS	Q03080 pteridium e
6	7	1.1	478	1 LXC2_PHOLE	P29236 photobacter
7	7	1.1	507	1 MS14_ARATH	O22607 arabidopsi
8	7	1.1	510	1 PUR1_CHICK	P28173 gallus gall
9	7	1.1	524	1 HEMK_RICCN	Q92913 rickettsia
10	7	1.1	535	1 TCER1_AYESA	P40412 avena sativ
11	7	1.1	535	1 TCER2_AYESA	P54411 avena sativ
12	7	1.1	535	1 TCPE_ARATH	O04450 arabidopsi
13	7	1.1	564	1 NOX1_HUMAN	Q9Y588 homo sapien
14	7	1.1	566	1 SCC1_YEAST	Q12358 saccharomy
15	7	1.1	579	1 MALT1_ARDAE	P13080 aedes aegypt
16	7	1.1	583	1 SCF3_SCHPO	P41000 schizosacch
17	7	1.1	585	1 ILV3_YEAST	P39522 saccharomy
18	7	1.1	591	1 UL49_EBV	P14347 Epstein-Barr
19	7	1.1	597	1 Y745_ARATH	P59278 arabidopsi
20	7	1.1	628	1 HNFA_MOUSE	P22561 mus musculu
21	7	1.1	628	1 HNFA_MOUSE	P15557 ratu
22	7	1.1	651	1 PDAT_YEAST	P40345 saccharomy
23	7	1.1	668	1 V244_FOWPV	Q9J426 fowlipox vir
24	7	1.1	700	1 MYBB_HUMAN	P10244 homo sapien
25	7	1.1	704	1 MYBB_MOUSE	P48972 mus musculu
26	7	1.1	731	1 BGAL_MALDO	P48981 malus domes
27	7	1.1	757	1 YHGF_NEIMA	P57072 neisseria m
28	7	1.1	757	1 YHGF_NEIMA	Q51152 neisseria m
29	7	1.1	757	1 YHGF_NEIMA	Q51152 neisseria m
30	7	1.1	843	1 NAL6_MOUSE	Q91W82 mus musculu
31	7	1.1	854	1 NAL6_MOUSE	Q63035 ratu
32	7	1.1	892	1 NAL6_MOUSE	P59044 homo sapien
33	7	1.1	938	1 TRG1_ECOLI	P33790 escherichia

34	7	1.1	945	1 AMPE_RAT	P50123 ratu
35	7	1.1	946	1 RHG4_HUMAN	P98171 homo sapien
36	7	1.1	966	1 MMBB_MYCTU	O53553 mycobacteri
37	7	1.1	1038	1 KF17_MOUSE	Q99PW8 mus musculu
38	7	1.1	1071	1 UBP7_YEAST	P40453 saccharomy
39	7	1.1	1099	1 CYA7_MOUSE	P51829 mus musculu
40	7	1.1	1163	1 LEPR_MACMU	Q9MY10 macaca mula
41	7	1.1	1165	1 LEPR_HUMAN	P48357 homo sapien
42	7	1.1	1168	1 DHX8_ARATH	Q38953 arabidopsi
43	7	1.1	1275	1 AIP1_MOUSE	Q9WVQ1 mus musculu
44	7	1.1	1277	1 AIP1_RAT	Q09804 schizosacch
45	7	1.1	1318	1 YAB2_SCHPO	Q20255 caenorhabdi
46	7	1.1	1332	1 YSL1_CAEEL	Q13428 homo sapien
47	7	1.1	1411	1 TCOF_HUMAN	Q86118 homo sapien
48	7	1.1	1455	1 AIP1_HUMAN	Q86118 homo sapien
49	7	1.1	1455	1 AIP1_HUMAN	Q86118 homo sapien
50	7	1.1	1572	1 MYHB_HUMAN	P35749 homo sapien
51	7	1.1	2077	1 TEGU_HSVG	P30002 human herpe
52	7	1.1	2077	1 TEGU_HSVG	P52340 human herpe
53	7	1.1	2210	1 RRPO_LYCAV	P14240 lymphocytic
54	7	1.1	2511	1 BPIE_MOUSE	Q91208 mus musculu
55	7	1.1	2768	1 THYG_HUMAN	P01266 homo sapien
56	7	1.1	3060	1 BPEB_HUMAN	Q8W8K8 homo sapien
57	7	1.1	3214	1 BPA1_HUMAN	Q03001 homo sapien
58	7	1.1	5038	1 PCLO_MOUSE	Q9GYX7 mus musculu
59	7	1.1	5085	1 PCLO_RAT	Q9JKE6 ratu
60	7	1.1	5147	1 PCLO_HUMAN	Q9Y5V0 homo sapien
61	7	1.1	5171	1 BPEA_HUMAN	Q94833 homo sapien
62	7	1.1	5262	1 MLT2_HUMAN	Q14686 homo sapien
63	7	1.1	7389	1 BPA1_MOUSE	Q91206 mus musculu
64	7	1.1	473	1 ATPB_BACSU	P37809 bacillus su
65	7	1.1	473	1 CTSY_SCHPO	P50214 schizosacch
66	7	1.1	473	1 IDH_ANASP	P50214 anabaena sp
67	7	1.1	473	1 VIL2_HPV16	P03107 human papil
68	7	1.1	473	1 ZF38_HUMAN	Q9Y5A6 homo sapien
69	7	1.1	475	1 DLDH_BUCBP	Q98948 buchnera ap
70	7	1.1	475	1 PAAB_ECOLI	P76083 escherichia
71	7	1.1	475	1 RMIC_ECOLI	P27850 escherichia
72	7	1.1	478	1 SEB4_HUMAN	O43236 homo sapien
73	7	1.1	478	1 SEB4_MOUSE	P28661 mus musculu
74	7	1.1	479	1 YP64_YEAST	Q12194 saccharomy
75	7	1.1	480	1 UCRI_BOVIN	P31000 bos taurus
			481	1 MENE_BACAA	Q81X97 bacillus an

ALIGNMENTS

RESULT 1	
DUSG_HUMAN	STANDARD; PRT; 665 AA.
AC	Q9BY84; Q9C0G3; Rel. 41, Created
DT	28-FEB-2003 (Rel. 41, Last sequence update)
DT	28-FEB-2003 (Rel. 41, Last sequence update)
DT	10-OCT-2003 (Rel. 42, Last annotation update)
DE	Dual specificity protein phosphatase 16 (EC 3.1.3.48) (EC 3.1.3.16)
DE	(Mitogen-activated protein kinase phosphatase 7) (MAP kinase phosphatase 7) (MKP-7).
DE	phosphatase 7) (MKP-7).
GN	DUSP16 OR MKP7 OR KIAA1700.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX	NCBI_TaxId=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=21486429; PubMed=11489891;
RA	Masuda K., Shima H., Watanabe M., Kikuchi K.;
RT	"MKP-7, a novel mitogen-activated protein kinase phosphatase, functions as a shuttle protein."
RL	J. Biol. Chem. 276:39002-39011(2001).
RP	[2]
RX	SEQUENCE FROM N.A.
RC	TISSUE=Brain;
RX	MEDLINE=21082932; PubMed=11214970;

RA Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XIX.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 7:347-355(2000).
 CC -1- FUNCTION: Involved in the inactivation of MAP kinases.
 CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H(2)O = protein
 CC tyrosine + phosphate.
 CC -1- CATALYTIC ACTIVITY: A phosphoprotein + H(2)O = a protein +
 CC phosphate.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic and nuclear.
 CC -1- SIMILARITY: Belongs to the protein-tyrosine phosphatase family.
 CC Non-receptor class dual specificity subfamily.
 CC -1- SIMILARITY: Contains 1 rhodanese domain.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 CC EMBL: AB052156; BAB40814.1; -;
 DR EMBL: AB051487; BAB2191.1; ALT_INT.
 DR HSSP: O16828; IMKP.
 DR Genew: HGNC:17909; DUSP16.
 DR MIM: 607175; -;
 DR GO: GO:0005737; C:cytoplasm; TAS.
 DR GO: GO:0005634; C:nucleus; TAS.
 DR GO: GO:0004721; F:protein phosphatase activity; TAS.
 DR GO: GO:0016311; F:dephosphorylation; TAS.
 DR GO: GO:0000188; P:inactivation of MAPK; TAS.
 DR GO: GO:0045209; P:leptomycin B-sensitive MAPK phosphatase nuc. .; TAS.
 DR GO: GO:0045204; P:MAPK nucleus export; TAS.
 DR InterPro: IPR000340; DS_phosphatase.
 DR InterPro: IPR008343; MAPK_phosph.
 DR InterPro: IPR001763; Rhodanese-like.
 DR InterPro: IPR000387; TYR_phosphatase.
 DR Pfam: PF00782; DSPC; 1.
 DR Pfam: PF00581; Rhodanese; 1.
 DR PRINTS: PR01764; MAPKPHPTASB.
 DR SMART: SM00195; DSPC; 1.
 DR SMART: SM00450; RHOD; 1.
 DR PROSITE: PS50206; RHODANES_3; 1.
 DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE: PS50056; TYR_PHOSPHATASE_2; 1.
 DR PROSITE: PS50054; TYR_PHOSPHATASE_DUAL; 1.
 KM Hydrolyase; Nuclear protein.
 FT DOMAIN 22 137 RHODANES.
 FT DOMAIN 227 289 PROTEIN-TYROSINE PHOSPHATASE.
 FT ACT_SITE 244 244 PHOSPHOCYSTEINE INTERMEDIATE (BY
 FT ACT_SITE 244 244 SIMILARITY)
 SQ SEQUENCE 665 AA; 73101 MW; 1BD853FF0846DDFF CRC64;
 Query Match 71.0%; Score 472; DB 1; Length 665;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 374 SPLVQALSGHLMSADRLSDSNKLRKSFSLDIKSYSAASMAASLHGSSSEDALRYTPS 433
 QY 434 TLLDGTINKLCQSPVOELSEQPTETSPDKEASIPKLIQTARPSPDSQSKRLHVSSTSSG 493
 DB 434 TLLDGTINKLCQSPVOELSEQPTETSPDKEASIPKLIQTARPSPDSQSKRLHVSSTSSG 493
 QY 494 TQKRLSLPRLHRSQGVENNYHTSPFLGISTGQHLTKSAGLGKGMHSDIILAPQSTPSL 553
 DB 494 TQKRLSLPRLHRSQGVENNYHTSPFLGISTGQHLTKSAGLGKGMHSDIILAPQSTPSL 553
 QY 554 TSSWTFATESSHFYASAIYGGASAYSCSQLPCTGQDYYSVRROKPSDRADSRSSW 613
 DB 554 TSSWTFATESSHFYASAIYGGASAYSCSQLPCTGQDYYSVRROKPSDRADSRSSW 613
 QY 614 HEESPFERQFRRSCQMEFGESIMENSRRELGVGSSQSSFGSMELIEVS 665
 DB 614 HEESPFERQFRRSCQMEFGESIMENSRRELGVGSSQSSFGSMELIEVS 665
 RESULT 2
 ID DUS8_HUMAN STANDARD; PRT; 625 AA.
 AC Q13202;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Dual specificity protein phosphatase 8 (EC 3.1.3.48) (EC 3.1.3.16)
 DE (Dual specificity protein phosphatase hvh-5).
 GN DUSP8 OR VH5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fetal brain;
 RX MEDLINE=96009533;
 RA Martell K.J., Seasholtz A.F., Kwak S.P., Clemens K.K., Dixon J.E.;
 RT "hvh-5, a protein tyrosine phosphatase abundant in brain that
 RT inactivates mitogen-activated protein kinase.";
 RL J. Neurochem. 65:1823-1833(1995).
 CC -1- FUNCTION: This protein shows both activity toward tyrosine-protein
 CC phosphatase as well as with serine/threonine-protein phosphatase (By
 CC similarity).
 CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H(2)O = protein
 CC tyrosine + phosphate.
 CC -1- CATALYTIC ACTIVITY: A phosphoprotein + H(2)O = a protein +
 CC phosphate.
 CC -1- TISSUE SPECIFICITY: Abundant in brain, heart and skeletal muscle.
 CC -1- SIMILARITY: Belongs to the protein-tyrosine phosphatase family.
 CC Non-receptor class dual specificity subfamily.
 CC -1- SIMILARITY: Contains 1 rhodanese domain.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 CC EMBL: U27193; AAA83151.1; -;
 DR HSSP: O16828; IMKP.
 DR Genew: HGNC:3074; DUSP8.
 DR MIM: 602038; -;
 DR GO: GO:0005737; C:cytoplasm; TAS.
 DR GO: GO:0004725; F:protein tyrosine phosphatase activity; TAS.
 DR GO: GO:0000188; P:inactivation of MAPK; TAS.
 DR GO: GO:0006470; P:protein amino acid dephosphorylation; TAS.
 DR InterPro: IPR000340; DS_phosphatase.
 DR InterPro: IPR008343; MAPK_phosph.
 DR InterPro: IPR001763; Rhodanese-like.
 DR InterPro: IPR000387; TYR_phosphatase.

```

DR Pfam; PF00782; DSPC; 1.
DR PRINTS; PR01764; MAPKPHPTASE.
DR SMART; SM00195; DSC; 1.
DR SMART; SM00450; RHOD; 1.
DR PROSITE; PS50206; RHODANSE_3; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS50054; TYR_PHOSPHATASE_DUAL; 1.
DR Hydrolyase; Nuclear protein.
KM DOMAIN 23 138 RHODANSE.
FT DOMAIN 162 430 PROTEIN-TYROSINE PHOSPHATASE.
FT DOMAIN 310 550 PRO-RICH.
FT ACT_SITE 246 246 PHOSPHOCYSTEINE INTERMEDIATE (BY
SQ SEQUENCE 625 AA; 65840 MW; DCBEA14487219666 CRC64;
SIMILARITY)).

Query Match 3.0%; Score 20; DB 1; Length 625;
Best Local Similarity 100.0%; Pred. No. 6.4e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 242 VHCLAGISRSATTAIAYIMK 261
Db 244 VHCLAGISRSATTAIAYIMK 263

RESULT 3
DUS8_MOUSE STANDARD; PRT; 663 AA.
ID DUS8_MOUSE
AC 009112;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Dual specificity protein phosphatase 8 (EC 3.1.3.48) (EC 3.1.3.16)
DE (Neuronal tyrosine threonine phosphatase 1).
DE DUSP8 OR NTRP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Scurionachti; Muridae; Murinae; Mus.
ON NCBI_TaxID=10090;
RX MEDLINE=69311565; PubMed=873137;
RA Theodosiou A.M., Rodrigues N.R., Nesbitt M.A., Ambrose H.J.,
RA Pateroson H., McElihan-Arnold E., Boyd Y., Leverisha M.A., Owen N.,
RA Blake D.J., Ashworth A., Davies K.E.;
RT "A member of the MAP kinase phosphatase gene family in mouse
RT containing a complex trinucleotide repeat in the coding region.";
RL Hum. Mol. Genet. 5:675-684(1996).
CC -1- FUNCTION: This protein shows both activity toward tyrosine-protein
CC phosphate as well as with serine/threonine-protein phosphate (By
CC similarity).
CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H(2)O = protein
CC tyrosine + phosphate.
CC -1- CATALYTIC ACTIVITY: A phosphoprotein + H(2)O = a protein +
CC phosphate.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic and nuclear.
CC -1- TISSUE SPECIFICITY: Expressed predominantly in brain and lung.
CC -1- SIMILARITY: Belongs to the protein-tyrosine phosphatase family.
CC Non-receptor class dual specificity subfamily.
CC -1- SIMILARITY: Contains 1 rhodanese domain.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; X95518; CA64772.1; -.
CC DR HSSP; Q16828; IMKP.
CC MGD; MGI:106626; Dusp8.

```

```

DR InterPro; IPR000340; DS_phosphatase.
DR InterPro; IPR008343; MAPK_phosph.
DR InterPro; IPR001763; Rhodanese-like.
DR InterPro; IPR000387; TYR_phosphatase.
DR Pfam; PF00782; DSPC; 1.
DR PRINTS; PR01764; MAPKPHPTASE.
DR SMART; SM00195; DSC; 1.
DR SMART; SM00450; RHOD; 1.
DR PROSITE; PS50206; RHODANSE_3; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS50054; TYR_PHOSPHATASE_DUAL; 1.
DR Hydrolyase; Nuclear protein.
KM DOMAIN 23 138 RHODANSE.
FT DOMAIN 162 432 POLY-ARG.
FT DOMAIN 452 459 POLY-SER.
FT DOMAIN 555 558 POLY-SER.
FT DOMAIN 559 576 POLY-GLY.
FT DOMAIN 577 600 POLY-SER.
FT DOMAIN 311 552 PRO-RICH.
FT ACT_SITE 246 246 PHOSPHOCYSTEINE INTERMEDIATE (BY
SQ SEQUENCE 663 AA; 68847 MW; 416F429A12C1FA7C CRC64;
SIMILARITY)).

Query Match 3.0%; Score 20; DB 1; Length 663;
Best Local Similarity 100.0%; Pred. No. 6.7e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 242 VHCLAGISRSATTAIAYIMK 261
Db 244 VHCLAGISRSATTAIAYIMK 263

RESULT 4
VHP1_CAEEL STANDARD; PRT; 619 AA.
ID VHP1_CAEEL
AC 010038;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Protein-tyrosine phosphatase vhp-1 (EC 3.1.3.48).
DE VHP-1 OR F08B1.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peldoderinae; Caenorhabditis.
ON NCBI_TaxID=6239;
RX MEDLINE=6239;
RA Chisoe S.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphatase + H(2)O = protein
CC tyrosine + phosphate.
CC -1- SIMILARITY: Belongs to the protein-tyrosine phosphatase family.
CC Non-receptor class dual specificity subfamily.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; U23178; AAC46719.1; -.
CC DR PIR; T15969; T15969.
CC DR HSSP; Q16828; IMKP.
CC DR WormPep; F08B1.1; CE01899.
CC DR InterPro; IPR000340; DS_phosphatase.
CC DR InterPro; IPR000387; TYR_phosphatase.
CC DR Pfam; PF00782; DSPC; 1.
CC DR SMART; SM00195; DSC; 1.
CC DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.

```

DR PROSITE; PSS0056; TYR_PHOSPHATASE_2; 1.
 DR PROSITE; PSS0054; TYR_PHOSPHATASE_DUAL; 1.
 KM Hydrolyase.
 FT DOMAIN 200 269 PROTEIN-TYROSINE PHOSPHATASE.
 FT DOMAIN 92 95 POLY-SER.
 FT DOMAIN 351 354 POLY-SER.
 FT DOMAIN 465 472 POLY-SER.
 FT DOMAIN 483 488 POLY-SER.
 FT ACT_SITE 224 224 PHOSPHOCYSTEINE INTERMEDIATE (BY SIMILARITY)
 SQ SEQUENCE 619 AA; 66354 MW; 3698326F615D0529 CRC64;
 Query Match 1.4%; Score 9; DB 1; Length 619;
 Best Local Similarity 100.0%; Pred. No. 1.2;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 243 HCLAGISRS 251
 Db 223 HCLAGISRS 231

RESULT 5
 ATP_PTRES STANDARD; PRT; 473 AA.
 AC 003080;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE ATP synthase beta chain (EC 3.6.3.14) (Fragment).
 GN ATPB.
 OS Pteridium esculentum.
 OG Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Moniliformes; Filicophyta; Filicopsida; Filicales;
 OC Dennstaedtiaceae; Pteridium.
 OX NCBI_TaxId=32102;
 RN [1]
 RS SEQUENCE FROM N.A.
 RA Wolf P.G.;
 RT "Evaluation of atpB nucleotide sequences for phylogenetic studies of ferns and other pteridophytes";
 RL Am. J. Bot. 84:1429-1440(1997).
 CC -1- FUNCTION: Produces ATP from ADP in the presence of a proton gradient across the membrane. The beta chain is the catalytic subunit.
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (in) = ADP + phosphate + H(+) (out).
 CC -1- SUBUNIT: F-type ATPases have 2 components, CF(1) - the catalytic core - and CF(0) - the membrane proton channel. CF(1) has five subunits: alpha(3), beta(3), gamma(1), delta(1), epsilon(1). CF(0) has three main subunits: a, b and c.
 CC -1- SUBCELLULAR LOCATION: Chloroplast thylakoid membrane.
 CC -1- SIMILARITY: Belongs to the ATPase alpha/beta chains family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U93834; AAB51742.1; -.
 DR InterPro; IPR003593; AAA_ATPase.
 DR InterPro; IPR005722; ATP_synthb1_beta.
 DR InterPro; IPR000793; ATPase_a/b_C.
 DR InterPro; IPR000194; ATPase_a/b_centre.
 DR InterPro; IPR004100; ATPase_a/b_N.
 DR InterPro; IPR009005; F1_ATPase_a/b_N.
 DR Pfam; PF00006; ATP-synt_ab; 1.
 DR Pfam; PF00306; ATP-synt_ab_C; 1.
 DR Pfam; PF02874; ATP-synt_ab_N; 1.
 DR SMART; SM00382; AAA; 1.

DR TIGRFAM; TIGR01039; atpD; 1.
 DR PROSITE; PS00152; ATPASE_ALPHA_BETA; 1.
 KM ATP synthase; Chloroplast; Thylakoid; Membrane; CF(1);
 KM Hydrolyase; Atp-binding; Hydrogen ion transport.
 FT NON_TER 1 1
 FT NP_BIND 172 179 ATP (POTENTIAL).
 FT NON_TER 473 473
 SQ SEQUENCE 473 AA; 50434 MW; 5F30596B88E90029 CRC64;
 Query Match 1.1%; Score 7; DB 1; Length 473;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 63 VLITELI 69
 Db 180 VLITELI 186

RESULT 6
 LXC2_PROHE STANDARD; PRT; 478 AA.
 AC P29236;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DE Acyl-CoA reductase (EC 1.2.1.50).
 GN LUXC.
 OS Photobacterium leiognathi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
 OC Vibrionaceae; Photobacterium.
 OX NCBI_TaxId=658;
 RN [1]
 RS SEQUENCE FROM N.A.
 RC STRAIN=ATCC 25521;
 RX MEDLINE=92007870; PubMed=1915359;
 RA Lee C.Y., Sziltner R.B., Meighen E.A.;
 RT "The lux genes of the luminous bacterial symbiont, Photobacterium leiognathi, of the ponyfish. Nucleotide sequence, difference in gene organization, and high expression in mutant Escherichia coli.";
 RL Eur. J. Biochem. 201:161-167(1991).
 CC -1- FUNCTION: LUXC IS THE FATTY ACID REDUCTASE ENZYME RESPONSIBLE FOR SYNTHESIS OF THE ALDEHYDE SUBSTRATE FOR THE LUMINESCENT REACTION CATALYZED BY LUCIFERASE.
 CC -1- CATALYTIC ACTIVITY: A long-chain aldehyde + CoA + NADPH = a long-chain acyl-CoA + NADP.
 CC -1- PATHWAY: Bioluminescent fatty acid reduction system; third step.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M63594; AAA25616.1; -.
 DR PIR; S17836; S17836.
 DR InterPro; IPR008670; LuxC.
 DR Pfam; PF05893; LuxC; 1.
 DR Luminance; Oxioreductase; NADP.
 SQ SEQUENCE 478 AA; 53713 MW; 4716F699BF7FE4A CRC64;

Query Match 1.1%; Score 7; DB 1; Length 478;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 SVHLIAG 123
 Db 145 SVHLIAG 151

RESULT 7
 MS14_ARATH

MS14 ARATH STANDARD; PRT: 507 AA.
 AC 022607; 042322; 042323; 093WF7; Q9SLD1;
 DT 15-UTL-1998 (Rel. 36, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE WD-40 repeat protein MS14.
 GN MS14 OR AT2G19520 OR F3P11.12.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eurosid II; Brassicales; Brassicaceae; Arabidopses.
 ON NCBI_TaxID=3702;
 RX MEDLINE=22954850; PubMed=14593172;
 RA Soutwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,
 RA Karlin-Nemman G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,
 RA Miranda-Memam G., Trip M., Chang C.H., Lee J.M., Tortum M.J.,
 RA Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,
 RA Arakawa T., Ban J., Banno F., Bowser L., Brookes S.Y., Carninci P.,
 RA Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,
 RA Hayaishi Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,
 RA Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,
 RA Kamita A., Meyers C., Nakajima M., Natsumura M., Seki M., Sakurai T.,
 RA Satou M., Tame R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,
 RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
 RT "Empirical analysis of transcriptional activity in the Arabidopsis
 RT genome.";
 RT Science 302:842-846(2003).
 RN [3]
 RP SEQUENCE OF 4-507 FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=99087420; PubMed=9872415;
 RA Kenzior A.L., Folk W.R.;
 RT "AtMS14 and RbAp48 WD-40 repeat proteins bind metal ions.";
 RN FEBS Lett. 440:425-429(1998).
 RN [4]
 RP SEQUENCE OF 1-142 AND 452-507 FROM N.A.
 RC STRAIN=cv. Columbia; TISSUE=Green silique;
 RA Raynal M., Grellet F., Landie M., Meyer Y., Cooke R., Delany M.;
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Nuclear (by similarity).
 CC -1- SIMILARITY: Belongs to the WD-repeat RBA46/RBA48/MS11 family.
 CC -1- CAUTION: Ref.4 (CAA85542) sequence differs from that shown due to
 CC a frameshift in position 136.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AY059799; AAL24281.1; -;
 DR EMBL; AY057655; AAL15286.1; -;
 DR EMBL; AY081447; AAM10009.1; -;
 DR EMBL; AF028711; AAD03340.1; -;
 DR EMBL; Z37286; CAA85542.1; ALT_FRAME.
 DR EMBL; Z37287; CAA85543.1; -;
 DR PIR; G84577; G84577.
 DR InterPro; IPR001580; WD40.
 DR Pfam; PF00400; WD40; 5.
 DR PRINTS; PR00320; GPROTEINRPT.
 DR ProDom; PD000018; WD40; 1.
 DR SMART; SM00320; WD40; 6.
 DR PROSITE; PS00678; WD_REPEATS_1; FALSE_NEG.
 DR PROSITE; PS50082; WD_REPEATS_2; 3
 DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
 KW Repeat; WD repeat; Nuclear protein.
 FT REPEAT 162 193 WD 1.
 FT REPEAT 217 248 WD 2.
 FT REPEAT 290 321 WD 3.
 FT REPEAT 335 366 WD 4.
 FT REPEAT 384 415 WD 5.
 FT REPEAT 439 477 WD 6.
 FT DOMAIN 468 471 POLY-GLY.
 FT CONFLICT 89 89 W -> L (IN REF. 4; CAA85542).
 FT CONFLICT 126 126 V -> F (IN REF. 4; CAA85542).
 FT CONFLICT 202 202 A -> P (IN REF. 3).
 FT CONFLICT 270 270 T -> P (IN REF. 3).
 FT CONFLICT 463 463 D -> A (IN REF. 4; CAA85543).
 FT CONFLICT 489 489 V -> F (IN REF. 4; CAA85543).
 FT CONFLICT 494 494 E -> A (IN REF. 3).
 SQ SEQUENCE 507 AA; 55759 MW; C37F8000F8B33397 CR664;
 Query Match 1.1%; Score 7; DB 1; Length 507;
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 302 TGAAGPK 308
 DB 20 TGAAGPK 26
 RESULT 8
 PIR1_CHICK STANDARD; PRT: 510 AA.
 AC P28173;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Amidophosphoribosyltransferase precursor (EC 2.4.2.14) (Glutamine
 DE phosphoribosylpyrophosphate amidotransferase) (ATPase) (GPAT).
 GN GPAT OR GPAT.
 OS Gallus gallus (Chicken).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 CC Gallus.
 ON NCBI_TaxID=9031;
 RX MEDLINE=91065928; PubMed=2123487;
 RA Zhou G., Dixon J.E., Zalkin H.;
 RL "Cloning and expression of avian glutamine
 RL phosphoribosylpyrophosphate amidotransferase. Conservation of a
 RL bacterial propeptide sequence supports a role for posttranslational
 RL processing.";
 RL J. Biol. Chem. 265:21152-21159(1990).
 RN [2]
 RP SEQUENCE OF 1-42 FROM N.A.
 RC MEDLINE=93330273; PubMed=8336716;
 RA Gavilanes A., Dixon J.E., Brayton K.A., Zalkin H.;
 RT "Coexpression of two closely linked avian genes for purine nucleotide
 RT synthesis from a bidirectional promoter.";
 RL Mol. Cell. Biol. 13:4784-4792(1993).
 RN [3]


```

RP ERRATUM.
RA Gavalas A., Dixon J.E., Brayton K.A., Zalkin H.;
RL Mol. Cell. Biol. 13:7977-7977(1993).
CC -1- CATALYTIC ACTIVITY: 5-phospho-beta-D-riboylamine + diphosphate +
CC L-glutamate = L-glutamine + 5-phospho-alpha-D-ribose 1-diphosphate
CC + H(2)O.
CC -1- COFACTOR: Binds 1 magnesium ion and 1 4Fe-4S cluster per subunit
CC (by similarity).
CC -1- SUBUNIT: De novo purine biosynthesis; first step.
CC -1- SIMILARITY: In the C-terminal section; belongs to the
CC purine/pyrimidine phosphoribosyltransferase family.
CC -1- SIMILARITY: Contains 1 type-2 glutamine amidotransferase domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M60063; AAA62736.1; -.
DR EMBL: L12533; AAA17895.1; -.
DR PIR: A38337; A38337.
DR HSP: P00497; 1A00.
DR MEROPS: C44.001; -.
DR InterPro: IPR005854; Amd_pshpho_trans.
DR InterPro: IPR000583; GATase_2.
DR InterPro: IPR002375; Pr_Py_Pd_transf.
DR InterPro: IPR000836; PRTtransferase.
DR Pfam: PF00310; GATase_2; 1.
DR Pfam: PF00156; Pribosyltran; 1.
DR TIGRFAMs: TIGR01134; purf_1.
DR PROSITE: PS00103; PUR_PVR_PR_TRANSFER; 1.
DR PROSITE: PS00443; GATASE_TYPE_II; 1.
DR Purine biosynthetas; Transferase; Glycosyltransferase;
DR Glutamine amidotransferase; Allosteric enzyme; Metal-binding;
DR Magnesium; Iron-sulfur; 4Fe-4S.
DR PROPEP 1 11 PROBABLE.
DR CHAIN 12 510 AMIDOPHOSPHORIBOSYLTRANSFERASE.
DR ACT SITE 12 12 GATASE (BY SIMILARITY).
DR METAL 280 280 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
DR METAL 327 327 MAGNESIUM (BY SIMILARITY).
DR METAL 389 389 MAGNESIUM (BY SIMILARITY).
DR METAL 390 390 MAGNESIUM (BY SIMILARITY).
DR METAL 426 426 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
DR METAL 496 496 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
DR METAL 499 499 IRON-SULFUR (4FE-4S) (BY SIMILARITY).
DR SEQUENCE 510 AA; 56257 MW; F4371FE1FEC7C744 CRC64;

Query Match 1.1%; Score 7; DB 1; Length 510;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 594 VYSVRRR 600
DB 299 VYSVRRR 305

RESULT 9
HEMK_RICCN STANDARD; PRT; 524 AA.
ID HEMK_RICCN
AC Q92613;
DT 10-OCT-2003 (Rel. 42; Created)
DT 10-OCT-2003 (Rel. 42; Last sequence update)
DT 10-OCT-2003 (Rel. 42; Last annotation update)
DE Bifunctional methyltransferase [includes: Hemk protein homolog
DE (EC 2.1.1.-) (M.RcoHemK) / tRNA (guanine-N(7)-)-methyltransferase
DE (EC 2.1.1.33) (tRNA(m7G46)-methyltransferase)].
OS HEMK OR RC1314.
OS Rickettsia conorii.
OS Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;

```

```

OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098(2001).
CC -1- FUNCTION: Catalyzes the formation of N(7)-methylguanine at
CC position 46 (m7G46) in tRNA (By similarity).
CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-
CC homocysteine + tRNA containing N(7)-methylguanine.
CC -1- SIMILARITY: In the N-terminal section; belongs to the hemk family.
CC -1- SIMILARITY: In the C-terminal section; belongs to the
CC methyltransferase superfamily. Trmb family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE008678; AAL03852.1; -.
DR HAMAP: MF_01057; fuesed; 1.
DR InterPro: IPR004395; Cons_hypoth91.
DR InterPro: IPR004556; Hemk_
DR InterPro: IPR003358; Methyltransf_4.
DR InterPro: IPR002052; NG_Mtase.
DR InterPro: IPR000051; SAM_bind.
DR Pfam: PF02390; Methyltransf_4; 1.
DR TIGRFAMs: TIGR00536; hemk fam; 1.
DR TIGRFAMs: TIGR00091; TIGR00091; 1.
DR PROSITE: PS00092; NG_MTASE; UNKNOWN 1.
DR Transferase; Methyltransferase; tRNA processing; Complete proteome.
FT DOMAIN 1 306 HEMK.
FT DOMAIN 307 524 TRNA (GUANINE-N(7)-)-METHYLTRANSFERASE.
FT SEQUENCE 524 AA; 60079 MW; 3E574DC6F7DFDC28 CRC64;

Query Match 1.1%; Score 7; DB 1; Length 524;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 178 ELIQONG 184
DB 473 ELIQONG 479

RESULT 10
TCEI_AVESA STANDARD; PRT; 535 AA.
ID TCEI_AVESA
AC P40412;
DT 01-FEB-1995 (Rel. 31; Created)
DT 01-FEB-1995 (Rel. 31; Last sequence update)
DT 01-OCT-1996 (Rel. 34; Last annotation update)
DE T-complex protein 1, epsilon subunit (TCP-1-epsilon) (CCT-epsilon)
DE (TCP-X19).
OS Avena sativa (Oat).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Avenae; Avena.
OX NCBI_TaxID=4498;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Dewi; TISSUE=coleoptile, and Mesocotyl;
RX MEDLINE=94085629; PubMed=7903257;
RA Elmann B., Krenz W., Munmerit E., Schaefer E.;
RA "Two Tcp-1-related but highly divergent gene families exist in oat
RT encoding proteins of assumed chaperone function.";

```

```

RL FEBS Lett. 336:313-316(1993).
CC -1- FUNCTION: Molecular chaperone; assist the folding of proteins upon
CC ATP hydrolysis. Known to play a role, in vitro, in the folding of
CC actin and tubulin.
CC -1- SUBUNIT: Hetero-oligomeric complex of about 850 to 900 kDa that
CC forms two stacked rings, 12 to 16 nm in diameter.
CC -1- SIMILARITY: Belongs to the TCP-1 chaperonin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X75777; CAA53396.1; -.
CC PIR; S40461; S40461.
CC HSSP; P48425; 1A6D.
CC InterPro; IPR002194; Chaperonin TCP-1.
CC InterPro; IPR001844; Chaperin Cpn60.
CC InterPro; IPR002423; Cpn60/TCP-1.
CC InterPro; IPR008950; GroEL-ATPase.
CC Pfam; PF00118; cpn60 TCP1; 1.
CC PRINTS; PR00298; CHAPERONIN60.
CC PROSITE; PS00750; TCP1_1; 1.
CC PROSITE; PS00751; TCP1_2; 1.
CC PROSITE; PS00995; TCP1_3; 1.
CC Chaperone; ATP-binding; Multigene family.
SQ SEQUENCE 535 AA; 58899 MW; 6A8F847CA891BC32 CRC64;

Query Match 1.1%; Score 7; DB 1; Length 535;
Best Local Similarity 100.0%; Pred. No. 1.le+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 KHKVDID 80
DB 259 KHKVDID 265

RESULT 11
TCPE2_AVEA ID STANDARD; PRT; 535 AA.
AC PS4411;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE T-complex protein 1, epsilon subunit (TCP-1-epsilon) (CCT-epsilon)
DE (TCP-K36).
OS Avena sativa (oat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poaceae;
OC Avenae; Avena.
OX NCBI_TaxID=498;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Pearl; TISSUE=coleoptile, and Mesocotyl;
RA MEDLINE=94085629; PubMed=7903257;
RA Eihman B., Krenz M., Nummert E., Schaefer E.;
RT "Two Tcp-1-related but highly divergent gene families exist in oat
RT encoding proteins of assumed chaperone function.";
RL FEBS Lett. 336:313-316(1993).
CC -1- FUNCTION: Molecular chaperone; assist the folding of proteins upon
CC ATP hydrolysis. Known to play a role, in vitro, in the folding of
CC actin and tubulin.
CC -1- SUBUNIT: Hetero-oligomeric complex of about 850 to 900 kDa that
CC forms two stacked rings, 12 to 16 nm in diameter.
CC -1- SIMILARITY: Belongs to the TCP-1 chaperonin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X75778; CAA53397.1; -.
CC PIR; S40462; S40462.
CC HSSP; P48424; 1A6D.
CC InterPro; IPR002194; Chaperonin TCP-1.
CC InterPro; IPR001844; Chaperin Cpn60.
CC InterPro; IPR002423; Cpn60/TCP-1.
CC InterPro; IPR008950; GroEL-ATPase.
CC Pfam; PF00118; cpn60 TCP1; 1.
CC PRINTS; PR00298; CHAPERONIN60.
CC PROSITE; PS00750; TCP1_1; 1.
CC PROSITE; PS00751; TCP1_2; 1.
CC PROSITE; PS00995; TCP1_3; 1.
CC Chaperone; ATP-binding; Multigene family.
SQ SEQUENCE 535 AA; 59034 MW; 02761226F5F1B81 CRC64;

Query Match 1.1%; Score 7; DB 1; Length 535;
Best Local Similarity 100.0%; Pred. No. 1.le+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 74 KHKVDID 80
DB 259 KHKVDID 265

RESULT 12
TCPE_ARATH ID STANDARD; PRT; 535 AA.
AC 004450;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE T-complex protein 1, epsilon subunit (TCP-1-epsilon) (CCT-epsilon).
DE AT1G24510 OR F21J9.12.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA MEDLINE=21016719; PubMed=1130712;
RA Theologis A., Becker J.R., Palm C.U., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Alcafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Eguu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizer L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin B.,
RA Kim C.U., Koo H.L., Kremetskaia I., Kutz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maltz R., Martelli A.,
RA Miltsecher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Tortum M.J., Town C.D.,
RA Uteback T., Van Aken S., Vayenberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Frazer C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana ";
RL Nature 408:816-820(2000).
CC -1- FUNCTION: Molecular chaperone; assist the folding of proteins upon
CC ATP hydrolysis. Known to play a role, in vitro, in the folding of
CC actin and tubulin.
CC -1- SUBUNIT: Hetero-oligomeric complex of about 850 to 900 kDa that

```

CC forms two stacked rings, 12 to 16 nm in diameter.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: Belongs to the TCP-1 chaperonin family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AC000103; AAF97977.1; -.
 DR HSSP: P48424; 1A6D.
 DR InterPro: IPR002194; Chaperonin_TCP-1.
 DR InterPro: IPR001844; Chaperin_Cpn60.
 DR InterPro: IPR002423; Cpn60/TCP-1.
 DR InterPro: IPR008950; GroEL-ATPase.
 DR Pfam: PF00118; cpn60_TCP1.1.
 DR PRINTS: PR00298; CHAPERONIN60.
 DR PRINTS: PR00304; TCOMPLEXTCP1.
 DR PROSITE: PS00750; TCP1_1; 1.
 DR PROSITE: PS00751; TCP1_2; 1.
 DR PROSITE: PS00995; TCP1_3; 1.
 DR Chaperone; ATP-binding; Multigene family.
 KW SEQUENCE 535 AA; 59384 MW; 1CB563433AFLD3 CRC64;
 QY 74 KHKVDID 80
 Db 259 KHKVDID 265
 RESULT 13
 NOXI_HUMAN
 ID NOXI_HUMAN STANDARD; PRT; 564 AA.
 AC Q9Y5S8; O95691;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE NADPH oxidase homolog 1 (NOX-1) (NOH-1) (NADH/NADPH mitogenic
 DE oxidase subunit P65-MOX) (Mitogenic oxidase 1) (MOX1).
 DE NOX1 OR NOH1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM NOH-1L).
 RA TISSUE=Colon epithelium;
 RX MEDLINE=99413719; PubMed=10485709;
 RA Suh Y.-A., Arnold R.S., Lassegue B., Shi J., Xu X., Sorrescu D.,
 RA Chung A.B., Griendling K.K., Lambeth J.D.;
 RT "Cell transformation by the superoxide-generating oxidase Mox1.";
 RL Nature 401:79-82(1999).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS NOH-1L, NOH-1LV AND NOH-1S).
 RA MEDLINE=20082959; PubMed=10615049;
 RA Banfi B., Maturana A., Jaconi S., Arnaudau S., Laforge T., Sinha B.,
 RA Ligeti E., Demareux N., Krause K.-H.;
 RT "A mammalian H+ channel, generated through alternative splicing of the
 RT NADPH oxidase homolog NOH-1.";
 RL Science 287:138-142(2000).
 RN [3]
 RP SEQUENCE OF 16-564 FROM N.A. (ISOFORM NOH-1L).
 RA Lloyd D.;
 RA Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: NOH-1S is a voltage-gated proton channel that mediates
 CC the H(+) currents of resting phagocytes and other tissues. It
 CC participates in the regulation of cellular pH and is blocked by

CC	zinc	NOH-1L is a pyridine nucleotide-dependent oxidoreductase
CC	that generates superoxide and might conduct H(+) ions as part of	
CC	its electron transport mechanism, whereas NOH-1S does not contain	
CC	an electron transport chain.	
CC	-1-COPFATOR: NADP and FAD (Potential).	
CC	-1-SUBCELLULAR LOCATION: Integral membrane protein (Potential).	
CC	-1-ALTERNATIVE PRODUCTS:	
CC	Event-Alternative splicing; Named isoforms=3;	
CC	Name=NOH-1L;	
CC	IsoId=Q9Y5S8-1; Sequence=Displayed;	
CC	Name=NOH-1S;	
CC	IsoId=Q9Y5S8-2; Sequence=VSP_001577, VSP_001578;	
CC	Name=NOH-1LV;	
CC	IsoId=Q9Y5S8-3; Sequence=VSP_001579;	
CC	-1-TISSUE SPECIFICITY: NOH-1L is detected in colon, uterus, prostate,	
CC	and colon carcinoma, but not in peripheral blood leukocytes. NOH-	
CC	1S is detected only in colon and colon carcinoma cells.	
CC	-1-SIMILARITY: Belongs to the FRE / CYBB family.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
DR	EMBL; AF127763; AAD38133.1; -	
DR	EMBL; AF166326; AAF23232.1; -	
DR	EMBL; AF166327; AAF23233.1; -	
DR	EMBL; AF166328; AAF23234.1; -	
DR	EMBL; Z83819; CAB06073.1; ALT_SEQ.	
DR	GeneW; HGNC:7889; NOX1.	
DR	MIM; 300225; -	
DR	GO; GO:0016021; C:integral to membrane; NAS.	
DR	GO; GO:0016175; P:superoxide-generating NADPH oxidase activity; TAS.	
DR	GO; GO:00301171; F:voltage-gated proton channel activity; TAS.	
DR	GO; GO:0006118; F:electron transport; NAS.	
DR	GO; GO:0006746; P:FADH2 metabolism; NAS.	
DR	GO; GO:0006739; P:NADPH metabolism; NAS.	
DR	GO; GO:0015992; P:proton transport; TAS.	
DR	InterPro; IPR002916; Ferric reduct.	
DR	InterPro; IPR000778; GP91phox.	
DR	Pfam; PF01794; Ferric_reduct; 1.	
DR	PRINTS; PR00466; GP91PHOX.	
KM	Oxidoreductase; NADP; Electron transport; Transmembrane; FAD; Heme;	
KM	Glycoprotein; Voltage-gated channel; Ionic channel;	
KM	Alternative splicing.	
FT	DOMAIN	1
FT	TRANSMEM	10
FT	POTENTIAL	30
FT	EXTRACELLULAR	31
FT	POTENTIAL	44
FT	TRANSMEM	45
FT	POTENTIAL	72
FT	CYTOSOLSMIC	73
FT	POTENTIAL	102
FT	TRANSMEM	103
FT	POTENTIAL	123
FT	EXTRACELLULAR	124
FT	POTENTIAL	168
FT	TRANSMEM	169
FT	POTENTIAL	189
FT	DOMAIN	190
FT	TRANSMEM	207
FT	POTENTIAL	227
FT	DOMAIN	228
FT	TRANSMEM	397
FT	POTENTIAL	417
FT	DOMAIN	418
FT	NP BIND	564
FT	METAL	338
FT	METAL	101
FT	METAL	101
FT	METAL	115
FT	METAL	115
FT	METAL	209
FT	METAL	221
FT	CARBOHYD	162
FT	CARBOHYD	236
FT	VASAPPLIC	159
FT	VASAPPLIC	190
FT	VARSPPLIC	191
FT	VARSPPLIC	564
FT	MISSING	(in isoform NOH-1S).
FT	/FTID=VSP_001577.	
FT	MISSING	(in isoform NOH-1S).

FT VARSPLIC 433 481 /FTId=VSP 001578.
 FT MISSING (In isoform NOH-1LV).
 FT CONFLICT 173 173 I -> V (IN REF. 2).
 SQ SEQUENCE 564 AA, 64870 MW, C3BE290F456DBC9A CRC64;
 Query Match 1.1%; Score 7; DB 1; Length 564;
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 644 BELGKVG 650
 DB 460 BELGKVG 466

RESULT 14
 SCCL1_YEAST STANDARD; PRT; 566 AA.
 AC 012158; 005325;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Sister chromatid cohesion protein 1.
 GN MCD1 OR SCCL1 OR RHC21 OR PDS3 OR YDL003W OR YD8119.04.
 OS Saccharomyces cerevisiae (Baker's Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OC NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.
 RC STRAIN=M303;
 RX MEDLINE=97474309; PubMed=9335333;
 RA Michaelis C., Clook R., Naemlyth K.;
 RT "Cohesins: chromosomal proteins that prevent premature separation of
 RT sister chromatids.";
 RL Cell 91:35-45(1997).
 RN [2]
 RP SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.
 RC STRAIN=S288C / FY1678;
 RX MEDLINE=97474310; PubMed=9335334;
 RA Guacci V., Koshland D., Strunnikov A.V.;
 RT "A direct link between sister chromatid cohesion and chromosome
 RT condensation revealed through the analysis of MCD1 in S. cerevisiae.";
 RL Cell 91:47-57(1997).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / FY1678;
 RX MEDLINE=97113263; PubMed=9169867;
 RA Jacq C., Alt-Moebie J., Andre B., Arnold W., Bahr A., Ballesta J.P.G.,
 RA Barues M., Baron L., Becker A., Biteau N., Bloecker H., Blugnon C.,
 RA Bogrovic T., Brandt P., Bruckner M., Bultrago M.J., Coster F.,
 RA Delvaux T., del Rey F., Dujon B., Eide L.G., Garcia-Cantalejo J.M.,
 RA Hoffeul A., Gomez-Petis A., Granocier C., Hanemann V., Hanken T.,
 RA Hoffeul J.D., Jaeger W., Jimenez A., Joniaux J.-L., Kraemer C.,
 RA Kueter H., Laamanen P., Legros Y., Louis B.J., Moeller-Rieker S.,
 RA Monnet A., Moro M., Mueller-Auer S., Nussbaumer B., Particio N.,
 RA Paulin L., Perez-Alonso M., Perez-Ortin J.E., Pohl T.M.,
 RA Prydz H., Purielle B., Rasmussen S.W., Remacha M., Revuelta J.L.,
 RA Rieger M., Salom D., Saluz H.P., Saiz J.E., Saren A.-M., Schaefer M.,
 RA Scherfe M., Schmidt E.R., Schneider C., Scholler P., Schwarz S.,
 RA Utratazu L.A., Verhasselt P., Vissers S., Voet M., Voickaert G.,
 RA Wagner G., Wandut R., Wedler E., Wedler H., Woelfl S., Harris D.E.,
 RA Bowman S., Brown D., Churcher C.M., Connor R., Dedman K., Gentles S.,
 RA Hamlin N., Hunt S., Jones L., McDonald S., Murphy L., Niblett D.,
 RA Odell C., Oliver K., Rajandream M.A., Richards C., Shore L.,
 RA Walsh S.V., Barrell B.G., Dietrich F.S., Mulligan J.T., Allen E.,
 RA Araujo R., Aviles E., Berno A., Carpenter J., Chen E., Cherry J.M.,
 RA Chung E., Duncan M., Hunnicke-Smith S., Hyman R.W., Komp C.,
 RA Lashkari D., Lew H., Lin D., Mosedale D., Nakahara K., Natch A.,
 RA Oelner P., Oh C., Petel F.X., Roberts D., Schramm S., Schroeder M.,
 RA Shogen T., Shroff N., Winant A., Yelton M.A., Botstein D.,
 RA Davis R.W., Johnson M., Andrews S., Brinkman R., Cooper J., Ding H.,
 Du Z., Favello A., Fulton L., Gatung S., Greco T., Hallsworth K.,

RA Hawkins J., Hillier L.W., Jier M., Johnson D., Johnston L.,
 RA Kirsten J., Kucaba T., Langston Y., Latreille P., Le T., Mardis E.,
 RA Meneses S., Miller N., Nhan M., Pauley A., Peluso D., Rikkin L.,
 RA Riles L., Tach A., Treaskis E., Vignati D., Wilcox L., Wolfdm P.,
 RA Vaudin M., Wais R., Waterston R., Albertmann K., Hant J., Heumann K.,
 RA Kleine K., Mewes H.-W., Zollner A., Zaccaria P.;
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome IV.";
 RL Nature 387:75-78(1997).
 RN [4]
 RP FUNCTION.
 RX MEDLINE=98151773; PubMed=9491073;
 RA Heo S.-J., Tatebayashi K., Kato J., Ikeda H.;
 RT "The RHC21 gene of budding yeast, a homologue of the fission yeast
 RT rad21+ gene, is essential for chromosome segregation.";
 RL Mol. Genet. 257:149-156(1998).
 RN [5]
 RP SUBCELLULAR LOCATION, INTERACTION WITH IRR1, IDENTIFICATION IN A
 RP COHESIN COMPLEX WITH SMC1, SMC3 AND IRR1, AND INTERACTION OF THE
 RP COHESIN COMPLEX WITH SCC2.
 RX MEDLINE=99145468; PubMed=9990856;
 RA Toch A., Clook R., Uhlmann F., Galova M., Schleiffer A., Naemlyth K.;
 RT "Yeast cohesin complex requires a conserved protein, Eco1p(Ctf17), to
 RT establish cohesion between sister chromatids during DNA replication.";
 RL Genes Dev. 13:320-333(1999).
 RN [6]
 RP CLEAVAGE BY ESPL1, FUNCTION, AND MUTAGENESIS OF ARG-180 AND ARG-268.
 RX MEDLINE=99330041; PubMed=10403247;
 RA Uhlmann F., Lotsepetch F., Naemlyth K.;
 RT "Sister-chromatid separation at anaphase onset is promoted by cleavage
 RT of the cohesin subunit Sccl1.";
 RL Nature 400:37-42(1999).
 RN [7]
 RP PHOSPHORYLATION BY CDC5, AND MUTAGENESIS OF SER-175 AND SER-263.
 RX MEDLINE=21264235; PubMed=11371343;
 RA Alexandru G., Uhlmann F., Mechtler K., Poudart M.-A., Naemlyth K.;
 RT "Phosphorylation of the cohesin subunit Sccl1 by Polo/Cdc5 kinase
 RT regulates sister chromatid separation in yeast.";
 RL Cell 105:459-472(2001).
 RN [8]
 RP IDENTIFICATION IN A COHESIN COMPLEX WITH SMC1, SMC3 AND IRR1, AND
 RP STRUCTURE.
 RX MEDLINE=21980168; PubMed=11983169;
 RA Haering C.H., Loewe J., Hochwagen A., Naemlyth K.;
 RT "Molecular architecture of SMC proteins and the yeast cohesin
 RT complex.";
 RL Mol. Cell 9:773-788(2002).
 CC -1- FUNCTION: Cleavable component of the cohesin complex involved in
 CC chromosome cohesion during cell cycle. The cohesin complex is
 CC required for the cohesion of sister chromatids after DNA
 CC replication. The cohesin complex apparently forms a large
 CC proteinaceous ring within which sister chromatids can be trapped.
 CC At metaphase-anaphase transition, this protein is cleaved by ESPL1
 CC and dissociates from chromatin, allowing sister chromatids to
 CC segregate.
 CC -1- SUBUNIT: Interacts directly with IRR1/SCC3 in cohesin complex.
 CC Cohesin complexes are composed of the SMC1 and SMC3 heterodimer
 CC attached via their hinge domain, MCD1/SCC1 which link them, and
 CC IRR1, which interacts with MCD1. The cohesin complex also
 CC interacts with SCC2, which is required for its association with
 CC chromosomes.
 CC -1- SUBCELLULAR LOCATION: Nuclear protein. Associates with chromatin.
 CC Before prophase it is scattered along chromosome arms. During
 CC prophase, most of cohesin complexes dissociate from chromatin
 CC except at centromeres, where cohesin complexes remain. At
 CC anaphase, it is cleaved by ESPL1, leading to the dissociation of
 CC the complex from chromosomes, allowing chromosome separation.
 CC -1- DOMAIN: The C-terminal part associates with the head of SMC1,
 CC while the N-terminal part binds to the head of SMC3.
 CC -1- PTM: Cleaved by ESPL1 at the onset of anaphase.
 CC -1- PTM: Phosphorylated by CDC5/Polo-like kinase at the onset of
 CC anaphase. Phosphorylation takes places at proximity to cleavage
 CC sites and is required for an efficient cleavage by ESPL1.
 CC -1- SIMILARITY: Belongs to the SCCL1/RAD21 family.

RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabbittowitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymoprez B.,
RA Welfens I., Vanstreels E., Rieger M., Schaefer M., Meiller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lebrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Medler H., Wambute R., Fumelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelure V., Mottier S.,
RA Gallbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado J., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armering J., Forzbach S.L.,
RA Cernetti L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,
RA Shpikoveki G.V., Useery D., Barrell B.G., Nurse P.,
RA "The genome sequence of *Schizosaccharomyces pombe*,"
RL Nature 415:871-880(2002).
RN [2]
RP SEQUENCE OF 108-583 FROM N.A.
RC STRAIN=972;
RX MEDLINE=95161095; PubMed=7857672;
RA Ishiguro J., Uhara Y., Kawahara K.;
RT "Molecular cloning and characterization of a fission yeast gene
RT responsible for supersensitivity to the spindle poison, isopropyl
RT N-3-chlorophenyl carbamate,"
RL Jpn. J. Genet. 69:671-678(1994).
RN [3]
RP REVISIONS.
RA Ishiguro J.;
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: Responsible for supersensitivity to the spindle poison,
CC isopropyl N-3-chlorophenyl carbamate.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 299260; CAB16391.1; -;
DR EMBL: AB017490; BAA33049.1; -;
DR PIR: T11624; T11624.
DR Genedb_SPombe; SPAC3A11.02; -;
DR InterPro; IPR000571; Znf.CCCH.
DR Pfam; PF00642; znf.CCCH; 2.
DR SMART; SM00356; ZNF_C3H1; 2.
SQ SEQUENCE 583 AA; 62807 MM; 905022C7C06B4271 CRC64;
QY
DB 408 SYSASNA 414
298 SYSASNA 304
RESULT 17
ID ILV3_YEAST STANDARD; PRT; 585 AA.
AC P39522;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dihydroxy-acid dehydratase, mitochondrial precursor (EC 4.2.1.9) (DAD)
DE (2,3-dihydroxy acid hydrolase).
GN ILV3 OR YOR016C OR J1450.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S28bc / FY1679;
RA de Haan M., Smits P.H.M., Grievell L.A.;
RL Submitted (MAY-1995) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE OF 82-585 FROM N.A.
RC STRAIN=X2180;
RX MEDLINE=94131281; PubMed=8299945;
RA Velasco J.A., Canasdo J., Pena M.C., Kawakami T., Laborda J.,
RA Notario V.;
RT "Cloning of the dihydroxyacid dehydratase-encoding gene (ILV3) from
RT Saccharomyces cerevisiae,"
RL Gene 137:179-185(1993).
CC -1- CATALYTIC ACTIVITY: 2,3-dihydroxy-3-methylbutanoate = 3-methyl-2-
CC oxobutanoate + H(2)O.
CC -1- COFACTOR: Binds 1 4Fe-4S cluster (Potential).
CC -1- PATHWAY: Valine and isoleucine biosynthesis; fourth step.
CC -1- SUBUNIT: Homodimer.
CC -1- SUBCELLULAR LOCATION: Mitochondrial.
CC -1- SIMILARITY: Belongs to the ILVD / edd family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X87611; CA60939.1; -;
DR EMBL: 249516; CAA89540.1; -;
DR EMBL: L13975; AAA34568.1; ALT_TERM.
DR PIR: S55205; S55205.
DR Genonline; 141853; -;
DR SGD: S000377; ILV3.
DR InterPro; IPR004404; ILVD.
DR InterPro; IPR000581; ILVD_EDD_family.
DR Pfam; PF00920; ILVD_EDD; 1.
DR Prodom; PD002691; ILVD_EDD_family; 1.
DR TIGRFAMs; TIGR00110; ILVD_1.
DR PROSITE; PS00886; ILVD_EDD_1.
DR PROSITE; PS00887; ILVD_EDD_2; 1.
DR TRANSIT peptide; Iron; Iron-sulfur; 4Fe-4S.
KW Branched-chain amino acid biosynthesis; Lyase; Mitochondrion;
KW TRANSIT peptide; Iron; Iron-sulfur; 4Fe-4S.
FT CHAIN 1 ? 585
FT METAL 143 143
FT METAL 221 221
FT METAL 82 89
FT CONFLICT 238 238
FT CONFLICT 242 242
FT CONFLICT 492 492
FT CONFLICT 520 520
FT CONFLICT 551 551
SQ SEQUENCE 585 AA; 62861 MM; 35AB3C679BA658D0 CRC64;
QY
DB 235 ASNGCVL 241
577 ASNGCVL 583
RESULT 18
ID ILV4_EBV STANDARD; PRT; 591 AA.
AC P14347;
DT 01-JAN-1990 (Rel. 13, Created)


```

DT 01-JAN-1990 (Rel. 13, Last sequence update)
DE 15-MAR-2004 (Rel. 43, Last annotation update)
DE Protein BFRF2.
GN BFRF2.
OC Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OC NCBI_TaxID=10377;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84270667; PubMed=6087149;
RA Beer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seglin C.,
RA Tufnell P.S., Barrett B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL Nature 310:207-211(1984).
CC -1- SIMILARITY: Belongs to the herpesviruses UL49 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; V01555; CA24880.1; ALT_INIT.
DR InterPro; IPR004339; UL49.
DR Pfam; PF03117; UL49; 1.
SQ SEQUENCE 591 AA; 63977 MW; EE63FF1E0721912E CRC64;
QY
DB 324 PAVSEGG 330
   |||||
   313 PAVSEGG 319

RESULT 19
Y745_ARATH STANDARD; PRT; 597 AA.
AC P59278; Q9C812;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Hypothetical protein Atg51745.
GN Atg51745 OR F19C24.5.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC euroids II; Brassicales; Brassicaceae; Arabidopsis.
OC NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altieri H., Araujo R., Bowman C.D., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Egtu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huzar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
RA Miltschev J., Miranda M., Nguyen M., Nieman W.C., Osborne B.I.,
RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,

```

```

RA Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana.";
RL Nature 406:816-820(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=22954850; PubMed=14593172;
RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,
RA Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,
RA Karlin-Newman G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,
RA Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.W., Toriumi M.J.,
RA Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari P.,
RA Arakawa T., Banh J., Banno F., Bowser L., Brooks S.Y., Carninci P.,
RA Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,
RA Hayashizaki Y., Johnson-Hopson C., Huan V.W., Iida K., Karnes M.,
RA Khan S., Koeseema E., Ishida J., Jiang P.X., Jones T., Kawai J.,
RA Kamya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,
RA Satou M., Tanase R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,
RA Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Empirical analysis of transcriptional activity in the Arabidopsis
RT genome.";
RL Science 302:842-846(2003).
CC -1- SIMILARITY: Contains 1 PWWP domain.
CC -1- CAUTION: Ref.1 sequence differs from that shown due to erroneous
CC gene model prediction.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AC025294; AAG50888.1; ALT_SEQ.
DR EMBL; BT002989; AAO2798.1; -.
DR InterPro; IPR000313; PWWP_domain.
DR PROSITE; PS50812; PWWP; 1.
FM Hypothetical protein.
FT DOMAIN 16
FT CONFICT 268 268 M -> R (IN REF. 1).
SQ SEQUENCE 597 AA; 65044 MW; E59141F5B9DFD79F CRC64;
QY
DB 230 IEXAKS 236
   |||||
   87 IEXAKS 93

RESULT 20
HNFA_MOUSE STANDARD; PRT; 628 AA.
ID HNFA_MOUSE
AC P22361;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hepatocyte nuclear factor 1-alpha (HNF-1A) (Liver-specific
DE transactivator factor 1F-B1) (TFB1).
GN TCF1 OR HNFA OR HNF-1A OR HNF-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91088607; PubMed=2263635;
RA Kuo C.U., Conley P.B., Hsieh C.L., Francke U., Crabtree G.R.;
RT "Molecular cloning, functional expression, and chromosomal

```


RT localization of mouse hepatocyte nuclear factor 1,"
 CC Proc. Natl. Acad. Sci. U.S.A. 87:9838-9842(1990).
 CC -1- FUNCTION: Required for the expression of several liver specific
 CC genes. Binds to the inverted palindrome 5'-GTTAATATTAC-3'.
 CC -1- SUBUNIT: Binds DNA as a dimer.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: Liver.
 CC -1- SIMILARITY: Belongs to the HNF1 homeobox family.
 CC -1- SIMILARITY: Contains 1 homeobox domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation on
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M57966; AAA37821.1; -
 CC PIR: A39262; A39262.
 CC PDB: 1F93; 20-SEP-00.
 CC PDB: 1G2Y; 17-JAN-01.
 CC PDB: 1G2Z; 17-JAN-01.
 CC PDB: 1G39; 17-JAN-01.
 CC PDB: 1J86; 11-JUL-01.
 CC PDB: 1LFB; 31-OCT-93.
 CC TRASPAC; T01211, -.
 CC MGD; MGI:98504; Tcf1.
 CC GO; GO:0005634; C:nucleus; IDA.
 CC InterPro; IPR006899; HNF-1_N.
 CC InterPro; IPR006898; HNF1A_C.
 CC InterPro; IPR006897; HNF1b_C.
 CC InterPro; IPR001356; Homeobox.
 CC Pfam; PF04814; HNF-1_N; 1.
 CC Pfam; PF04813; HNF-1A_C; 1.
 CC Pfam; PF04812; HNF-1B_C; 1.
 CC SMART; SMO0389; HOX; 1.
 CC PROSITE; PS00027; HOMEBOX_1; 1.
 CC PROSITE; PS00027; HOMEBOX_2; 1.
 CC Transcription regulation; DNA-binding; Homeobox; Nuclear protein;
 CC Activator; Trans-acting factor; 3D-structure.
 CC K1 DOMAIN 1 80 DIMERIZATION.
 CC K2 DOMAIN 1 31 ASP/GLU-RICH (ACIDIC) (POTENTIAL
 CC INVOLVEMENT WITH TRANSCRIPTION).
 CC K3 DOMAIN 1 80 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 CC FT DOMAIN 197 205
 CC FT DNA BIND 199 219
 CC FT DOMAIN 238 258
 CC FT 21 AMINO ACID LOOP BETWEEN HELIX 2 AND 3.
 CC SQ SEQUENCE 628 AA; 67237 MW; 737920D1A36B9DD CRC64;
 CC -----
 CC Query Match 1.1%; Score 7; DB 1; Length 628;
 CC Best Local Similarity 100.0%; Pred. No. 1.3e+02;
 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 CC -----
 CC QY 344 SATSEAA 350
 CC |||||
 CC DB 325 SATSEAA 331
 CC -----
 CC RESULT 21
 CC HNF1A_RAT
 CC ID HNF1A_RAT STANDARD; PRT; 628 AA.
 CC AC P15257;
 CC DT 01-APR-1990 (Rel. 14, Created)
 CC DT 01-APR-1990 (Rel. 14, Last sequence update)
 CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
 CC DE Hepatocyte nuclear factor 1-alpha (HNF-1A) (Liver-specific
 CC transcription factor Lf-B1) (LFB1).
 CC GN TCFL1 OR HNF1A OR HNF-1A OR HNF-1.
 CC OS Rattus norvegicus (Rat).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 CC OX NCBI_TaxId=10116;
 CC RN [1]

RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=90003224; PubMed=2571419;
 RA Frain M., Swart G., Monaci P., Nicotia A., Staempfli S., Frank R.,
 RA Cortese R.;
 RT "The liver-specific transcription factor Lf-B1 contains a highly
 RT diverged homeobox DNA binding domain.";
 RL Cell 59:145-157(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91016926; PubMed=2216777;
 RA Chouard T., Blumenfeld M., Bach I., Vandekerckhove J., Cereghini S.,
 RA Yaniv M.;
 RT "A distal dimerization domain is essential for DNA-binding by the
 RT atypical HNF1 homeodomain.";
 RL Nucleic Acids Res. 18:5853-5863(1990).
 RN [3]
 RP SEQUENCE OF 166-628 FROM N.A.
 RX MEDLINE=90249741; PubMed=1970973;
 RA Baumhueter S., Mendel D.B., Conley P.B., Kuo C.J., Turk C.,
 RA Graves M.K., Edwards C.A., Courtois G., Crabtree G.R.;
 RT "HNF-1 shares three sequence motifs with the POU domain proteins and
 RT is identical to Lf-B1 and Apf.";
 RN Genes Dev. 4:372-379(1990).
 RN [4]
 RP SEQUENCE OF 1-12 FROM N.A.
 RC STRAIN=Sprague-Dawley;
 RA Tomei L., Piaggio G., Tonietti C., Lazzerio D., de Francesco R.,
 RA Pozzi L., Gerstner J., Cortese R.;
 RL Submitted (AUG-1992) to the EMBL/Genbank/DBJ databases.
 RN [5]
 RP POSITION OF HOMEBOX.
 RX MEDLINE=90106643; PubMed=1967225;
 RA Finney M.;
 RT "The homeodomain of the transcription factor Lf-B1 has a 21 amino
 RT acid loop between helix 2 and helix 3.";
 RL Cell 60:5-6(1990).
 RN [6]
 RP STRUCTURE BY NMR OF 1-32.
 RX MEDLINE=91105074; PubMed=1988016;
 RA Pastore A., de Francesco R., Barbato G., Castiglione Morelli M.A.,
 RA Motta A., Cortese R.;
 RT "1H resonance assignment and secondary structure determination of the
 RT dimerization domain of transcription factor LfB1.";
 RL Biochemistry 30:148-153(1991).
 RN [7]
 RP STRUCTURE BY NMR OF 195-286.
 RX MEDLINE=93259120; PubMed=8491172;
 RA Leitling B., de Francesco R., Tomei L., Cortese R., Otting G.,
 RA Wuehrich K.;
 RT "The three-dimensional NMR-solution structure of the polypeptide
 RT fragment 195-286 of the LfB1/HNF1 transcription factor from rat liver
 RT comprises a nonclassical homeodomain.";
 RL EMBO J. 12:1797-1803(1993).
 RN [8]
 RP STRUCTURE BY NMR OF 195-286.
 RX MEDLINE=97272000; PubMed=9126845;
 RA Schott O., Billeter M., Leitling B., Wider G., Wuehrich K.;
 RT "The NMR solution structure of the non-classical homeodomain from the
 RT rat liver LfB1/HNF1 transcription factor.";
 RL J. Mol. Biol. 267:673-683(1997).
 RN [9]
 RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 195-286.
 RX MEDLINE=93259121; PubMed=8491173;
 RA Ceska T.A., Lamers M., Monaci P., Nicotia A., Cortese R., Suck D.;
 RT "The X-ray structure of an atypical homeodomain present in the rat
 RT liver transcription factor LfB1/HNF1 and implications for DNA
 RT binding.";
 RL EMBO J. 12:1805-1810(1993).
 RN [10]
 RP FUNCTION: Required for the expression of several liver specific
 CC genes. Binds to the inverted palindrome 5'-GTTAATATTAC-3'.
 CC -1- SUBUNIT: Binds DNA as a dimer.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC RN [11]

```

CC -1- TISSUE SPECIFICITY: Liver.
CC -1- SIMILARITY: Belongs to the HNF1 homeobox family.
CC -1- SIMILARITY: Contains 1 homeobox domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, J03170; AAA41524.1; -
DR EMBL, X54423; CAA38295.1; -
DR EMBL, X67649; CAA47891.1; -
DR EMBL, X53297; CAA37387.1; ALT_INIT.
DR PIR, A33333; A33333.
DR PIR, S25485; S25485.
DR PDB, 1LFB; 31-OCT-93.
DR PDB, 2LFB; 12-MAR-97.
DR TRANSFAC, T00369; -
DR InterPro, IPR006889; HNF-1_N.
DR InterPro, IPR006898; HNF1A_C.
DR InterPro, IPR006897; HNF1B_C.
DR InterPro, IPR001356; Homeobox.
DR Pfam, PF04814; HNF-1_N; 1.
DR Pfam, PF04813; HNF-1A_C; 1.
DR Pfam, PF04812; HNF-1B_C; 1.
DR SMART, SM00389; HOX; 1.
DR PROSITE, PS00027; HOMEBOX_1; 1.
DR PROSITE, PS50071; HOMEBOX_2; 1.
DR Transcription regulation; DNA-binding; Homeobox; Nuclear protein;
DR Activator; Trans-acting factor; 3D-structure.
CC KW DOMAIN 1 31 DIMERIZATION.
CC FT DOMAIN 1 80 ASP/GLU-RICH (ACIDIC) (POTENTIAL
CC FT 197 205 INVOLVEMENT WITH TRANSCRIPTION).
CC FT 199 279 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
CC FT 238 258 HOMEBOX.
CC FT 208 218 21 AMINO ACID LOOP BETWEEN HELIX 2 AND 3.
CC FT 219 220
CC FT 226 241
CC FT 242 245
CC FT 248 249
CC FT 251 254
CC FT 255 257
CC FT 261 273
CC FT 274 275
CC SQ SEQUENCE 628 AA; 67213 MW; 8028099308C86A52 CRC64;

Query Match 1.1%; Score 7; DB 1; Length 628;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 344 SATSEEA 350
DB 325 SATSEEA 331

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=S28bc / FY1679;
RA MEDLINE=95208356; PubMed=7900425;
RA Verlaesselt P., Aert R., Voet M., Volckaert G.;
RT "Twelve open reading frames revealed in the 23.6 kb segment flanking
RT the centromere on the Saccharomyces cerevisiae chromosome XIV right
RT arm."
RT Yeast 10:1355-1361(1994).
RN (2)
RN CHARACTERIZATION.
RX MEDLINE=20300927; PubMed=10829075;
RA Dahlqvist A., Stahl U., Lenman M., Banas A., Lee M., Sandager L.,
RA Rønne H., Styrene S.;
RT "Phospholipid:diacylglycerol acyltransferase: an enzyme that catalyzes
RT the acyl-CoA-independent formation of triacylglycerol in yeast and
RT plants."
RL Proc. Natl. Acad. Sci. U.S.A. 97:6487-6492(2000).
CC -1- FUNCTION: Triacylglycerol formation by an acyl-CoA independent
CC pathway. The enzyme specifically transfers acyl groups from the
CC sn-2 position of a phospholipid to diacylglycerol, thus forming
CC an sn-1-lysophospholipid.
CC -1- CATALYTIC ACTIVITY: Phospholipid + 1,2-diacylglycerol =
CC lysophospholipid + triacylglycerol.
CC -1- SIMILARITY: SOME, TO MAMMALIAN PHOSPHATIDYLCHOLINE-STEROL O-
CC ACYLTRANSFERASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, X77395; CAA54576.1; -
DR EMBL, Z71623; CAA96285.1; -
DR PIR, S45131; S45131.
DR GenBank, U14353; -
DR SGD, S0005291; LRO1.
DR GO, GO:0046027; F:Phospholipid:diacylglycerol acyltransferase. .; IDA.
DR GO, GO:0019915; P:Lipid storage; IDA.
DR GO, GO:0019432; P:triacylglycerol biosynthesis; IDA.
DR InterPro, IPR003385; LACT.
DR Pfam, PF02450; LACT; 1.
DR Transferrase; Acyltransferase; Transmembrane.
FT TRANSMEM 81 101 POTENTIAL.
FT SEQUENCE 661 AA; 75393 MW; 01C043319A836F44 CRC64;

Query Match 1.1%; Score 7; DB 1; Length 661;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 421 SSSSDAL 427
DB 431 SSSSDAL 437

```

```

RESULT 22
PDAT_YEAST STANDARD; PRT; 661 AA.
AC P40345;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Phospholipid:diacylglycerol acyltransferase (EC 2.3.1.158) (PDAT).
GN LRO1 OR YNR008W OR N2042.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]

```

```

RESULT 23
V244_FOWPV STANDARD; PRT; 668 AA.
AC O9J426;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative ankryrin-repeat protein FPV244.
GN FPV244.
OS Fowlpox virus (FPV).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Avipoxvirus.
OX NCBI_TaxID=10261;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20193820; PubMed=10729156;

```

RA Alfonso C.L., Tulman E.R., Lu Z., Zaak L., Kutish G.F., Rock D.L.;
 RT "The genome of fowlpox virus";
 CC J. Virol. 74:3815-3831(2000).
 CC -1- SIMILARITY: Contains 12 ANK repeats.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF198100; AAF4588.1; -
 DR InterPro: IPR002110; ANK.
 DR Pfam: PF00023; ank; 10.
 DR SMART: SM00248; ANK; 11.
 DR PROSITE: PSS0088; ANK_REPEAT; 7.
 DR PROSITE: PSS0297; ANK_REPEAT_REGION; 1.
 KW Hypothetical protein; Repeat; ANK repeat.
 FT REPEAT
 FT 40 69 ANK 1.
 FT REPEAT 144 173 ANK 2.
 FT REPEAT 177 206 ANK 3.
 FT REPEAT 210 239 ANK 4.
 FT REPEAT 272 302 ANK 5.
 FT REPEAT 306 336 ANK 6.
 FT REPEAT 340 370 ANK 7.
 FT REPEAT 374 403 ANK 8.
 FT REPEAT 407 437 ANK 9.
 FT REPEAT 441 471 ANK 10.
 FT REPEAT 473 502 ANK 11.
 FT REPEAT 571 602 ANK 12.
 SQ SEQUENCE 668 AA; 76200 MW; F3C5BA076023791B CRC64;
 QY 177 KELLQON 183
 Db 58 KELLQON 64
 Query Match 1.1%; Score 7; DB 1; Length 668;
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
 RA Coulson A., Cowille G.J., Deadman R., Dhani P.D., Dunn M.,
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 RA Hammond D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
 RA Hargrave S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 RA Leivaeslino M.H., Leverhwa M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 RA Marsh V.U., Martin S.L., McCormachie L.J., McIlroy J.C., Nickerson T.,
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
 RA Piliavinore B.J.C.T., Prachalangam S.R., Plumb R.W., Ramsey H.,
 RA Rice C.M., Rose M.T., Scott C.E., Sehra H.K., Showmreen R., Sims S.,
 RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 RA Tracey A., Tromans A.C., Vaudin M., Walli M., Wallis J.M.,
 RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 RA Rogers J.;
 RT "The DNA sequence and comparative analysis of human chromosome 20.";
 RL Nature 414:865-871(2001).
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain, and Eye;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Bueltow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F.,
 RA Datchenko L., Marusina K., Farmer A.A., Rubin G.W., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant P., Scheetz T.E.,
 RA Brownstein M.J., Uebli T.B., Toshitsuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.J., Peters G.J., Abramson R.D., Mullaly S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Aleksov J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren K., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalka U., Smalke D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [4]
 RP PHOSPHORYLATION OF THR-444; THR-487; THR-494 AND SER-577.
 RX MEDLINE=99054657; PubMed=9840932;
 RA Saville M.K., Watson R.U.;
 RT "The cell-cycle regulated transcription factor B-Myb is phosphorylated
 RT by cyclin A/Cdk2 at sites that enhance its transactivation
 RT properties.";
 RL Oncogene 17:2679-2689(1998).
 RN [5]
 RP PHOSPHORYLATION OF THR-440; THR-444; THR-494 AND SER-577.
 RX MEDLINE=99195476; PubMed=10095772;
 RA Bartsch O., Horstmann S., Toprak K., Klempner K.H., Ferrari S.;
 RT "Identification of cyclin A/Cdk2 phosphorylation sites in B-Myb.";
 RL Eur. J. Biochem. 260:384-391(1999).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- PTM: PHOSPHORYLATED BY CYCLIN A/CDK2 DURING S-PHASE.
 CC PHOSPHORYLATION AT THR-520 IS PROBABLY INVOLVED IN TRANSCRIPTIONAL
 CC ACTIVITY.
 CC -1- SIMILARITY: Contains 3 Myb-like domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

```

CC -----
DR EMBL; AL13293; CAA31655.1; -
DR EMBL; X121866; CAC08192.1; -
DR EMBL; BC007585; AA07585.1; -
DR EMBL; BC053555; AA03355.1; -
DR PIR; S01991; S01991.
DR HSSP; Q03237; 1A5J.
DR TRANSFAC; T00065; -
DR Genew; HGNC:7548; MYBL2.
DR GK; P10244; -
DR MIM; 601415; -
DR GO; GO:0003700; F:transcription factor activity; TAS.
DR InterPro; IPR001005; Myb DNA binding.
DR Pfam; PF00249; myb DNA-binding; 3.
DR SMART; SM00717; SANT; 3.
DR PROSITE; PS00037; MYB_1; 3.
DR PROSITE; PS00334; MYB_2; 3.
DR PROSITE; PS50090; MYB_3; 3.
DR Transcription regulation; Nuclear protein; DNA-binding; Repeat;
Phosphorylation.
KW DNA_BIND 26 77 MYB 1.
FT DNA_BIND 78 129 MYB 2.
FT MOD_RES 130 180 MYB 3.
FT MOD_RES 440 440 PHOSPHORYLATION (BY CDK2).
FT MOD_RES 444 444 PHOSPHORYLATION (BY CDK2).
FT MOD_RES 487 487 PHOSPHORYLATION (BY CDK2).
FT MOD_RES 494 494 PHOSPHORYLATION (BY CDK2).
FT MOD_RES 520 520 PHOSPHORYLATION (BY CDK2).
FT MOD_RES 577 577 PHOSPHORYLATION (BY CDK2).
SQ SEQUENCE 700 AA; 78764 MW; D81B28B3DAB94061 CRC64;

Query Match 1.1%; Score 7; DB 1; Length 700;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 CSQKVV 87
DB 479 CSQKVV 485

RESULT 25
MYB_MOUSE
ID MYB_MOUSE STANDARD; PRT; 704 AA.
AC P48972;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myb-related protein B (B-Myb).
GN MYBL2 OR BMYB.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9236176; PubMed=1501895;
RA Lam E.W., Robinson C., Watson R.J.;
RT "Characterization and cell cycle-regulated expression of mouse
RT B-myb."
RL Oncogene 7:1685-1690(1992).
RP SEQUENCE FROM N.A.
RC TISSUE=Limb;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Lonnellano N.A., Peters G.J., Abramson R.D., Millar J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

```

```

RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Vyllalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Whiting J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield V.S.N., Krzywinski M.I., Skalek U., Smalls D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Maita M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE OF 1-6 FROM N.A.
RX MEDLINE=93327760; PubMed=8334989;
RA Lam E.W., Watson R.J.;
RT "An E2F-binding site mediates cell-cycle regulated repression of
RT mouse B-myb transcription."
RL EMBO J. 12:2705-2713(1993).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- PTM: PHOSPHORYLATED BY CYCLIN A/CDK2 DURING S-PHASE.
CC PHOSPHORYLATION AT THR-524 IS PROBABLY INVOLVED IN TRANSCRIPTIONAL
CC ACTIVITY (BY SIMILARITY).
CC -1- SIMILARITY: Contains 3 Myb-like domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X70472; CAA49898.1; -
DR EMBL; BC050842; AA050842.1; -
DR EMBL; X73028; CAA51511.1; -
DR PIR; S33704; S33704.
DR HSSP; Q03237; 1A5J.
DR MGD; MGI:101785; Myb12.
DR InterPro; IPR001005; Myb DNA binding.
DR Pfam; PF00249; myb DNA-binding; 3.
DR SMART; SM00717; SANT; 3.
DR PROSITE; PS00037; MYB_1; 3.
DR PROSITE; PS00334; MYB_2; 3.
DR PROSITE; PS50090; MYB_3; 3.
KW Transcription regulation; Nuclear protein; DNA-binding; Repeat;
Phosphorylation.
FT DNA_BIND 26 77 MYB 1.
FT DNA_BIND 78 129 MYB 2.
FT DNA_BIND 130 180 MYB 3.
FT MOD_RES 443 443 PHOSPHORYLATION (BY CDK2) (BY
FT MOD_RES 447 447 SIMILARITY).
FT MOD_RES 490 490 PHOSPHORYLATION (BY CDK2) (BY
FT MOD_RES 497 497 SIMILARITY).
FT MOD_RES 524 524 PHOSPHORYLATION (BY CDK2) (BY
FT MOD_RES 581 581 SIMILARITY).
SQ SEQUENCE 704 AA; 79102 MW; 0EF09C1E2184E47 CRC64;

```

```

Query Match 1.1%; Score 7; DB 1; Length 704;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 CSQKVV 87
DB 482 CSQKVV 488

RESULT 26

```

```

BGAL_MALDO
ID BGAL_MALDO STANDARD; PRT; 731 AA.
AC P48981;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 26-FEB-2003 (Rel. 41, Last annotation update)
DE Beta-galactosidase precursor (EC 3.2.1.23) (Lactase) (Acid beta-
galactosidase) (Exo-(1-->4)-beta-D-galactanase).
OS Malus domestica (Apple) (Malus sylvestris)
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC Eurosid1; Rosales; Rosaceae; Maloideae; Malus.
OX NCBI_TaxID=3750;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RC STRAIN=cv. Granny Smith; TISSUE=fruit cortical tissue;
RA MEDLINE=95083752; PubMed=791682;
RA Ross G.S., Weigryn T., Macrae E.A., Redgwell R.J.;
RT "Apple beta-galactosidase. Activity against cell wall polysaccharides
RT and characterization of a related cDNA clone.";
RL Plant Physiol. 106:521-528(1994).
CC -1- FUNCTION: Involved in cell wall degradation. Degrades
CC polysaccharides containing beta-(1-->4)-linked galactans, acting
CC as an exo-(1-->4)-beta-D-galactanase.
CC -1- CATALYTIC ACTIVITY: Hydrolysis of terminal, non-reducing beta-D-
CC galactose residues in beta-D-galactosides.
CC -1- SIMILARITY: Belongs to family 35 of glycosyl hydrolases.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; L29451; AAA62324.1; -.
DR PIR; T17002; T17002.
DR InterPro; IPR008979; Gal_bind_1like.
DR InterPro; IPR001944; Glyco_hydro_35.
DR Pfam; PF01301; Glyco_hydro_35; 1.
DR PRINTS; PR00742; GLYHYDRLASE35.
DR PROSITE; PS01182; GLYCOSYL_HYDROL_F35; 1.
KM Hydrolase; Glycosidase; Signal.
FT SIGNAL 1 23
FT CHAIN 24 731
FT ACT_SITE 182 182
FT ACT_SITE 251 251
FT ACT_SITE NUCLEOPHILE (POTENTIAL).
SQ SEQUENCE 731 AA; 80995 MW; FAB65D24A0D30BD4 CRC64;

Query Match 1.1%; Score 7; DB 1; Length 731;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 648 KVSQSS 654
Db 414 KVSQSS 420

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=22491 / Serogroup A / Serotype 4A;
RA MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltham T., Hamlin N., Holtroyd S.,
RA Jorgels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrall B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis 22491.";
RL Nature 404:502-506(2000).
CC -1- SIMILARITY: Contains 1 SI motif domain.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AL162752; CAB83508.1; -.
DR PIR; E82013; E82013.
DR HSP; P05055; ISRO.
DR InterPro; IPR008994; Nucleic_acid_OB.
DR InterPro; IPR003029; SI.
DR InterPro; IPR006641; YqgFc.
DR Pfam; PF00575; SI; 1.
DR SMART; SM00316; SI; 1.
DR SMART; SM00732; YqgFc; 1.
DR PROSITE; PS50126; SI; 1.
KM Hypothetical protein; RNA-binding; Complete proteome.
FT DOMAIN 640 709
FT DOMAIN SI MOTIF.
FT CONFLICT 265 266
SQ SEQUENCE 757 AA; 83161 MW; C7E9731BD4BC0F3 CRC64;

Query Match 1.1%; Score 7; DB 1; Length 757;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 638 SENSRE 644
Db 723 SENSRE 729

```

```

RESULT 27
YHGF_NEIMA
ID YHGF_NEIMA STANDARD; PRT; 757 AA.
AC P57072;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein NMA0194.
GN NMA0194.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=5699;
RN [1]

```

```

RESULT 28
YHGF_NEIMB
ID YHGF_NEIMB STANDARD; PRT; 757 AA.
AC G51152;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein NMB0075.
GN NMB0075.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B1940 / Serogroup B;
RA MEDLINE=96236055; PubMed=8655518;
RA Peterling H., Hammerschmidt S., Froesch M., van Putten J.P.M.,
RA Ison C.A., Robertson B.D.;
RT "Genes associated with meningococcal capsule complex are also found
RT in Neisseria gonorrhoeae";
RL J. Bacteriol. 178:3342-3345(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RA MEDLINE=20175755; PubMed=10710307;
RA Tetzelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,

```

```

RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwin M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Macon T., Ciecko A., Perkey D.S., Blair E., Ciltone H., Clark E.B.,
RA Coton M.D., Uterback T.R., Khouri H., Qin H., Yamathavan J.,
RA Gill J., Scariato V., Massiani V., Piza M., Grandi G., Sun L.,
RA Smith H.O., Frazer C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58";
RL Science 287:1809-1815(2000).
CC -1- SIMILARITY: Contains 1 S1 motif domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L09189; AAC37046.1; -.
DR EMBL; AE002366; AAF40542.1; -.
DR PIR; A81242; A81242.
DR HSP; P05055; JSRO.
DR TIGR; NMB0075; -.
DR InterPro; IPR008994; Nucleic_acid_OB.
DR InterPro; IPR003029; S1.
DR Pfam; PF00575; S1; 1.
DR SMART; SM00316; S1; 1.
DR SMART; SM00732; YqgFc; 1.
DR PROSITE; PS50126; S1; 1.
DR Hypothetical protein; RNA-binding; Complete proteome.
FT DOMAIN 640 709 S1 MOTIF
FT CONFLICT 265 266 WL -> CV (IN REF. 1).
SQ SEQUENCE 757 AA; 83161 MW; 7B6851EB5ED6AE CRC64;

Query Match 1.1%; Score 7; DB 1; Length 757;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 638 SENSRE 644
DB 723 SENSRE 729

RESULT 29
DEGY CAEEL STANDARD; PRT; 795 AA.
AC 001635;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Degenerin-like protein ZK770.1 in chromosome I.
GN ZK770.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Chordata; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bricol N2.
RA Wagerl L., Gattung S., Bartko L.;
RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: Belongs to the amiloride-sensitive sodium channel
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U97404; AAB33309.1; -.
DR PIR; T34468; T34468.
DR WormBep; ZK770.1; CE15411.
DR InterPro; IPR004726; Deg-1.
DR InterPro; IPR001873; Na+channel_ASC.
DR Pfam; PF00858; ASC; 1.
DR PRINTS; PR01078; AMINACHANNEL.
DR TRIPFAME; TIGR00867; deg-1; 1.
DR PROSITE; PS01206; ASC; 1.
KW Hypothetical protein; Ionic channel; Transmembrane; Ion transport;
KW Glycoprotein; Sodium channel.
FT DOMAIN 1 58 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 59 79 POTENTIAL.
FT DOMAIN 80 739 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 740 760 POTENTIAL.
FT DOMAIN 761 795 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 267 269 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 367 367 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 388 388 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 458 458 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 499 499 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 518 518 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 795 AA; 90776 MW; 2CAAC7D41770B54 CRC64;

Query Match 1.1%; Score 7; DB 1; Length 795;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 406 SVSYSAS 412
DB 665 SVSYSAS 671

RESULT 30
NAL6 MOUSE
ID NAL6 MOUSE STANDARD; PRT; 843 AA.
AC Q91W52; Q8K014;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE PYRIN-containing APAF1-like protein 5-like.
GN NALP6 OR PYPAF5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP CONCEPTUAL TRANSLATION OF 1-287.
RA Hinz U.;
RL Unpublished observations (FEB-2003).
RN [2]
RP SEQUENCE OF 288-843 FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heist F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toehiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loggiano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richardson S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A.C., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butlerfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,

```

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [3]
 RP IDENTIFICATION OF MAMMALIAN ORTHOLOGS OF PYPAF5.
 RX MEDLINE=22520874; PubMed=12633874;
 RA Albrecht M., Domingues F.S., Schreiber S., Lengauer T.;
 RT "Identification of mammalian orthologs associates PYPAF5 with distinct
 functional roles";
 RL FEBS Lett. 538:173-177(2003).
 CC -1- FUNCTION: May mediate activation of CASP1 via ASC and promote
 CC activation of NF-kappa-B (By similarity).
 CC -1- SUBUNIT: Binds to ASC with its DAPIN domain (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: Contains 1 DAPIN domain.
 CC -1- SIMILARITY: Contains 1 NACHT domain.
 CC -1- SIMILARITY: Contains 3 leucine-rich (LRR) repeats.
 CC -1- CAUTION: The N-terminus was extended using ESTs and genomic
 CC sequences, in analogy to ortholog sequences.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; BC013519; AAH13519.1; -;
 DR EMBL; BC031139; AAH31139.1; ALT_INIT.
 DR MGD; MGI:2141990; Pypaf5.
 DR InterPro; IPR001611; LRR.
 DR InterPro; IPR007091; LRR_RNinh.
 DR InterPro; IPR007111; NACHT_NTPase.
 DR InterPro; IPR004020; PAAD_DAPIN_dom.
 DR Pfam; PF00560; LRR; 1.
 DR Pfam; PF05729; NACHT; 1.
 DR Pfam; PF02758; PAAD_DAPIN; 1.
 DR PROSITE; PS50824; DAPIN; 1.
 DR PROSITE; PS50837; NACHT; 1.
 DR ATP-binding; Leucine-rich repeat; Repeat.
 DR DOMAIN; 11 102 DAPIN.
 FT DOMAIN 168 485 NACHT.
 FT DOMAIN 569 585 ASP/GU-RICH.
 FT DOMAIN 654 661 POLY-LYS.
 FT REPEAT 434 459 LRR 1.
 FT REPEAT 609 632 LRR 2.
 FT REPEAT 811 834 LRR 3.
 FT NP BIND 174 181 ATP (POTENTIAL).
 SO SEQUENCE 843 AA; 94592 MW; 35FB7A766A47DB51 CRC64;

Query Match 1.1%; Score 7; DB 1; Length 843;
 Best Local Similarity 100.0%; Pred. 1.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 270 AYRFVKE 276
 DB 343 AYRFVKE 349

RESULT 31
 ID NAL6_RAT STANDARD; PRT; 854 AA.
 AC 063035;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE PRIN-containing APAF1-like protein 5-like (Angiotensin II/vasopressin
 DE receptor).
 GN NALP6 OR PYPAF5 OR AVR.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RP CONCEPTUAL TRANSLATION OF 1-343.
 RA Hinz U.;
 RL Unpublished observations (FEB-2003).
 RN [2]
 RP SEQUENCE OF 344-854 FROM N.A., MUTAGENESIS OF GUU-393, AND FUNCTION.
 RX MEDLINE=96071640; PubMed=7489366;
 RA Ruiz-Opazo N., Akimoto K., Herrera V.L.M.;
 RT "Identification of a novel dual angiotensin II/vasopressin receptor on
 RT the basis of molecular recognition theory";
 RL Nat. Med. 1:1074-1081(1995).
 RN [3]
 RP VARIANTS SALT-SENSITIVE HYPERTENSION SER-492 AND ARG-536.
 RX MEDLINE=21980185; PubMed=11984003;
 RA Ruiz-Opazo N., Lopez L.V., Herrera V.L.M.;
 RT "The dual AngII/AVP receptor gene N119S/C163R variant exhibits
 RT sodium-induced dysfunction and cosegregates with salt-sensitive
 RT hypertension in the Dahl salt-sensitive hypertensive rat model";
 RL Mol. Med. 8:24-32(2002).
 RN [4]
 RP IDENTIFICATION OF MAMMALIAN ORTHOLOGS OF PYPAF5.
 RX MEDLINE=22520874; PubMed=12633874;
 RA Albrecht M., Domingues F.S., Schreiber S., Lengauer T.;
 RT "Identification of mammalian orthologs associates PYPAF5 with distinct
 RT functional roles";
 RL FEBS Lett. 538:173-177(2003).
 CC -1- FUNCTION: May mediate activation of CASP1 via ASC and promote
 CC activation of NF-kappa-B (By similarity). Angiotensin II and
 CC vasopressin binding protein. May stimulate cAMP accumulation.
 CC -1- SUBUNIT: Binds to ASC with its DAPIN domain (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- TISSUE SPECIFICITY: Highly expressed in kidney. Detected at low
 CC levels in all tissues tested.
 CC -1- DISEASE: Defects in NALP6 may be a cause of salt-sensitive
 CC hypertension.
 CC -1- SIMILARITY: Contains 1 DAPIN domain.
 CC -1- SIMILARITY: Contains 1 NACHT domain.
 CC -1- SIMILARITY: Contains 4 leucine-rich (LRR) repeats.
 CC -1- CAUTION: The N-terminus was extended using ESTs and genomic
 CC sequences, in analogy to ortholog sequences.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M85183; AAA03623.1; ALT_INIT.
 DR InterPro; IPR001611; LRR.
 DR InterPro; IPR007091; LRR_RNinh.
 DR InterPro; IPR007111; NACHT_NTPase.
 DR InterPro; IPR004020; PAAD_DAPIN_dom.
 DR Pfam; PF00560; LRR; 2.
 DR Pfam; PF05729; NACHT; 1.
 DR Pfam; PF02758; PAAD_DAPIN; 1.
 DR PROSITE; PS50824; DAPIN; 1.
 DR PROSITE; PS50837; NACHT; 1.
 DR ATP-binding; Leucine-rich repeat; Repeat.
 DR DOMAIN; 1 103 DAPIN.
 FT DOMAIN 168 484 NACHT.
 FT DOMAIN 564 604 ASP/GU-RICH.
 FT DOMAIN 655 662 POLY-LYS.
 FT REPEAT 433 458 LRR 1.
 FT REPEAT 610 633 LRR 2.
 FT REPEAT 722 745 LRR 3.
 FT REPEAT 812 836 LRR 4.
 FT NP BIND 174 181 ATP (POTENTIAL).
 FT VARIANT 492 492 N -> S (IN SALT-SENSITIVE HYPERTENSION).


```

FT VARIANT 536 536 C -> R (IN SALT-SENSITIVE HYPERTENSION)
FT MTJAGEN 766 766 E->K; ABOLISHES ANGIOTENSIN II BINDING.
SQ SEQUENCE 854 AA; 95292 MW; D7BBD922D77B3734 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 854;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 270 AYPFKE 276
Db 343 AYPFKE 349

RESULT 32
NAL6_HUMAN
ID NAL6_HUMAN STANDARD; PRT; 892 AA.
AC P59044;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE NACHT, LRR- and PYD-containing protein 6 (PYRIN-containing APAF1-like
DE protein 5)
GN NALP6 OR PYPAF5.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
[1]
RX MEDLINE=22162427; PubMed=12019269;
RA Wang L., Manji G.A., Grenier J.M., Al-Garawi A., Merriam S.,
RA Lora J.M., Geddes B.J., Briskin M., Distefano P.S., Bertin J.;
RT "PYPAF7, a novel PYRIN-containing Apaf1-like protein that regulates
RT activation of NF-kappa B and caspase-1-dependent cytokine
RT processing.";
RL J. Biol. Chem. 277:29674-29680(2002).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22451042; PubMed=12563287;
RA Tschopp J., Martinon F., Burns K.;
RT "NALP6: a novel protein family involved in inflammation.";
RL Nat. Rev. Mol. Cell Biol. 4:95-104(2003).
RN [3]
RP FUNCTION.
RX MEDLINE=22275822; PubMed=12387869;
RA Grenier J.M., Wang L., Manji G.A., Huang W.-J., Al-Garawi A.,
RA Kelly R., Carlson A., Merriam S., Lora J.M., Briskin M.,
RA Distefano P.S., Bertin J.;
RT "Functional screening of five PYPAF family members identifies PYPAFs
RT as a novel regulator of NF-kappaB and caspase-1.";
RL FEBS Lett. 530:73-78(2002).
CC -1- FUNCTION: May mediate activation of CASP1 via ASC and promote
CC activation of NF-kappa-B.
CC -1- SUBUNIT: Binds to ASC with its DAPIN domain.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: Highly expressed in granulocytes. Detected at
CC much lower levels in T-cells.
CC -1- SIMILARITY: Contains 1 DAPIN domain.
CC -1- SIMILARITY: Contains 1 NACHT domain.
CC -1- SIMILARITY: Contains 5 leucine-rich (LRR) repeats.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, AF479748; AAL87105.1; -.
CC EMBL, AY154461; AAO18157.1; -.
CC Genew: HGNC:22944; NALP6.
CC InterPro: IPR000767; Disease_resist.

```

```

DR InterPro: IPR007091; LRR_RNinh.
DR InterPro: IPR003590; LRR_RNinh_sub.
DR InterPro: IPR007111; NACHT_NTPase.
DR InterPro: IPR004020; PAAD_DAPIN_dom.
DR Pfam: PF05729; NACHT; 1.
DR Pfam: PF02758; PAAD_DAPIN; 1.
DR PRINTS: PR00364; DISFASERISIT.
DR SMART: SM00368; LRR_R1; 3.
DR PROSITE: PS00824; DAPIN; 1.
DR PROSITE: PS00837; NACHT; 1.
KW ATP-binding; Leucine-rich repeat; Repeat.
FT DOMAIN 1 103 DAPIN.
FT DOMAIN 196 513 NACHT.
FT DOMAIN 604 614 POLY-GLU.
FT REPEAT 462 487 LRR 1.
FT REPEAT 727 747 LRR 2.
FT REPEAT 755 778 LRR 3.
FT REPEAT 811 834 LRR 4.
FT REPEAT 845 868 LRR 5.
FT NP_BIND 202 209 ATP (POTENTIAL).
SQ SEQUENCE 892 AA; 98733 MW; 4AABDFC766DDE9D CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 1; Length 892;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 270 AYPFKE 276
Db 371 AYPFKE 377

RESULT 33
TRG1_ECOLI
ID TRG1_ECOLI STANDARD; PRT; 938 AA.
AC P33790;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Trg protein.
GN TRG.
OS Escherichia coli.
OG Plasmid F.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxId=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=K12;
RC MEDLINE=92204127; PubMed=1348105;
RA Firth N., Skurray R.A.;
RT "Characterization of the F plasmid bifunctional conjugation gene,
RT trgG.";
RL Mol. Gen. Genet. 232:145-153(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94359430; PubMed=7915817;
RA Frost L.S., Ippen-Ihler K., Skurray R.A.;
RT "Analysis of the sequence and gene products of the transfer region of
RT the F sex factor.";
RL Microbiol. Rev. 58:162-210(1994).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / CR63;
RA Shimizu H., Satoh Y., Suda Y., Uehara K., Sampaio G., Mizobuchi K.;
RT "Complete nucleotide sequence of the F plasmid: its implications for
RT organization and diversification of plasmid genomes.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PLAYS A CRUCIAL ROLE IN DONOR-RECIPIENT CELL
CC INTERACTIONS. REQUIRED FOR TWO STAGES OF THE CONJUGATION PROCESS:
CC PLUS BIOSYNTHESIS AND MATING AGGREGATE STABILIZATION. MAY
CC INTERACT WITH TRAN.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC -1- PTM: TRAG* PROBABLY ARISES FROM THE POST-TRANSLATIONAL CLEAVAGE OF

```

TRAG
-!- CAUTION: TRAG IS NOT RESPONSIBLE FOR THE N-TERMINAL ACETYLATION
OF F PLIN AS STATED BY SOME AUTHORS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC DR EMBL, MS9763; AAA98081.1; -.
CC DR EMBL, U01159; AAC44184.1; -.
CC DR EMBL, AP001918; BAA97969.1; -.
CC DR PIR, S20480; S20480.
CC DR Ecogene; EG40114; trag.
CC Plasmid; conjugation; Transmembrane; Inner membrane;
CC Complete proteome.
CC FT CHAIN 1 938 TRAG PROTEIN.
CC FT CHAIN ? 938 TRAG PROTEIN.
CC FT DOMAIN 1 53 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 54 73 POTENTIAL.
CC FT DOMAIN 74 329 PERIPLASMIC (POTENTIAL).
CC FT TRANSMEM 330 350 POTENTIAL.
CC FT DOMAIN 351 412 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 413 433 POTENTIAL.
CC FT DOMAIN 434 938 PERIPLASMIC (POTENTIAL).
CC SQ SEQUENCE 938 AA; 102471 MW; 980908402ACGPDF CRC64;

Query Match 1.1%; Score 7; DB 1; Length 938;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 417 LHGFSS 423
Db 544 LHGFSS 550

RESULT 34
AMPE RAT
ID AMPE RAT STANDARD; PRT; 945 AA.
AC P50123; Q64200; Q9ULQ7; Q9ULQ9;
DT 01-OCT-1996 (Rel. 34, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Gluamyl aminopeptidase (EC 3.4.11.7) (EAP) (Aminopeptidase A) (APA).
GN ENPEP.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND TISSUE SPECIFICITY.
RX STRAIN=Sprague-Dawley; TISSUE=hippocampus;
RC MEDLINE=20435312; PubMed=10978538;
RA Lee H.-J., Tomioka M., Takaki Y., Masumoto H., Saido T.C.;
RT "Molecular cloning and expression of aminopeptidase A isoforms from
rat hippocampus.";
RL Biochim. Biophys. Acta 1493:273-278(2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.
RC TISSUE=Kidney;
RX MEDLINE=2031186; PubMed=10666143;
RA Troyanovskaya M., Jayaraman G., Song L., Healy D.P.;
RT "Aminopeptidase-A. I. cDNA cloning and expression and localization in
rat tissues.";
RL Am. J. Physiol. 278:R413-R424(2000).
RN [3]
RP SEQUENCE OF 265-397 FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=95029920; PubMed=7943354;
RA Song L., Ye M., Troyanovskaya M., Wilk E., Wilk S., Healy D.P.;

```

RT "Rat kidney glumamyl aminopeptidase (aminopeptidase A): molecular
RT identity and cellular localization.",
RL Am. J. Physiol. 267:F546-F557(1994).
RN [4]
RP SEQUENCE OF 482-606 FROM N.A. (ISOFORM 1), AND TISSUE SPECIFICITY.
RC TISSUE=Kidney;
RX MEDLINE=96188975; PubMed=8613196;
RA Troyanovskaya M., Song L., Jayaraman G., Healy D.P.;
RT "Expression of aminopeptidase A, an angiotensinase, in glomerular
RT mesangial cells.";
RL Hypertension 27:518-522(1996) .
CC -I- FUNCTION: Appears to have a role in the catabolic pathway of the
CC renin-angiotensin system. Isoform 1 has aminopeptidase activity
CC while isoform 2 does not.
CC -I- CATALYTIC ACTIVITY: Release of a N-terminal glutamate (and to a
CC lesser extent aspartate) from a peptide.
CC -I- COFACTOR: Binds 1 zinc ion per subunit (By similarity) .
CC -I- SUBUNIT: Homodimer; disulfide-linked.
CC -I- SUBCELLULAR LOCATION: Type II membrane protein.
CC -I- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1; Synonyms=APAL;
CC IsoId=P50123-1; Sequence=Displayed;
CC Name=2; Synonyms=APAS;
CC IsoId=P50123-2; Sequence=VSP_007845;
CC CC -I- TISSUE SPECIFICITY: Highest expression in kidney proximal tubules
CC and ileum enterocytes. High expression also detected in liver and
CC placental. Lower levels in heart, adrenal gland and brain. Not
CC detected in aorta, lung or spleen. In heart, higher levels in
CC ventricle than in atrium. Also expressed in glomerular mesangial
CC cells.
CC -I- SIMILARITY: Belongs to peptidase family M1.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.jab-sib.ch/announce/
CC or send an email to license@jab-sib.ch).
-----
DR EMBL; AF146044; AAPE6704.1; -.
DR EMBL; AF146518; AAP66710.1; -.
DR EMBL; AF214568; AAP37622.1; -.
DR MEROPS; M01.003; -.
DR InterPro; IPR006025; Pept_M_Zn_BS.
DR InterPro; IPR001930; Peptidase_M1.
DR Pfam; PF01433; Peptidase_M1; 1.
DR PRINTS; PR00756; ALMADIPTASE.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
KW Hydrolase; Metalloprotease; Aminopeptidase; Zinc; Calcium;
KW Glycoprotein; Transmembrane; Phosphorylation; Signal-anchor;
KW Alternative splicing.
FT DOMAIN 1 17 CYTOPLASMIC (POTENTIAL).
FT TRANSEM 18 40 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT TRANSSEM 18 40 (POTENTIAL).
FT DOMAIN 41 945 EXTREMECTERIAL (POTENTIAL).
FT METAL 385 385 ZINC (CATALYTIC) (BY SIMILARITY) .
FT ACT_SITE 386 386 BY SIMILARITY..
FT METAL 389 389 ZINC (CATALYTIC) (BY SIMILARITY) .
FT METAL 408 408 ZINC (CATALYTIC) (BY SIMILARITY) .
FT ACT_SITE 471 471 PROTON DONOR (POTENTIAL) .
FT MOD_RES 12 12 PHOSPHORYLATION (BY SIMILARITY) .
FT CARBOHD 116 116 N-LINKED (GLCNAC. . . ) (POTENTIAL) .
FT CARBOHD 189 189 N-LINKED (GLCNAC. . . ) (POTENTIAL) .
FT CARBOHD 236 236 N-LINKED (GLCNAC. . . ) (POTENTIAL) .
FT CARBOHD 316 316 N-LINKED (GLCNAC. . . ) (POTENTIAL) .
FT CARBOHD 546 546 N-LINKED (GLCNAC. . . ) (POTENTIAL) .
FT CARBOHD 584 584 N-LINKED (GLCNAC. . . ) (POTENTIAL) .
FT CARBOHD 601 601 N-LINKED (GLCNAC. . . ) (POTENTIAL) .
FT CARBOHD 640 640 N-LINKED (GLCNAC. . . ) (POTENTIAL) .
FT CARBOHD 669 669 N-LINKED (GLCNAC. . . ) (POTENTIAL) .
FT CARBOHD 754 754 N-LINKED (GLCNAC. . . ) (POTENTIAL) .

```

```

FT CARBOHYD 766 766 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 792 792 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 569 573 YTMNI -> NHREC (in isoform 2).
FT VARSPLIC 574 945 /FTID=VSP_007844.
FT VARSPLIC 574 945 Missing (in isoform 2).
FT CONFLICT 256 256 K -> E (IN REF. 1).
FT CONFLICT 306 306 N -> K (IN REF. 1).
FT CONFLICT 748 748 A -> D (IN REF. 1).
SQ SEQUENCE 945 AA; 107994 MM; 5A74F1A537DC537 CRC64;

Query Match 1.1%; Score 7; DB 1; Length 945;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 290 QLLDYER 296
DB 660 QLLDYER 666

RESULT 35
RHGA_HUMAN STANDARD; PRT; 946 AA.
ID P98171; Q14144;
AC 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Rho-GAP-activating protein 4 (Rho-GAP hematopoietic protein C1)
DE (p115).
GN ARHGAP4 OR RHOGAP4 OR RGCI OR KIAA0131.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA TRIBIO1 C., Dreotto S., Bione S., Cesareni G., Torrisi M.R.,
RA Lotti L.V., Lanfrancone L., Toniolo D., Pelicci P.;
RA "An X chromosome-linked gene encoding a protein with characteristics
RT of a rhoGAP predominantly expressed in hematopoietic cells.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:695-699(1996).
RN [2]
RP SEQUENCE OF 5-946 FROM N.A.
RX MEDLINE=96127530; PubMed=8590280;
RA Nagase T., Seki N., Tanaka A., Ishikawa K.-I., Momura N.;
RA "Prediction of the coding sequences of unidentified human genes. IV.
RT The coding sequences of 40 new genes (K1A0121-K1A0160) deduced by
RL DNA Res. 2:167-174(1995).
RN [3]
RP SEQUENCE OF 1-103 FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=95072568; PubMed=7981673;
RA Tribio1 C., Mancini M., Plassart B., Bione S., Rivella S.,
RA Sala C., Torri G., Toniolo D.;
RA "Isolation of new genes in distal Xq28: transcriptional map and
RT identification of a human homologue of the ARD1 N-acetyl transferase
RL of Saccharomyces cerevisiae.";
Hum Mol. Genet. 3:1061-1068(1994).
CC -1- FUNCTION: Inhibitory effect on stress fiber organization. May
CC down-regulate Rho-like GTPase in hematopoietic cells.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC, JUST BELOW THE PLASMA MEMBRANE.
CC -1- TISSUE SPECIFICITY: Predominantly in hematopoietic cells (spleen,
CC thymus and leukocytes); low levels in placenta, lung and various
CC fetal tissues.
CC -1- SIMILARITY: Contains 1 FCH domain.
CC -1- SIMILARITY: Contains 1 Rho-GAP domain.
CC -1- SIMILARITY: Contains 1 SH3 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X78817; CAA55394.1; -
CC EMBL; D50921; BAA09480.1; -
CC PIR; I38100; I38100.
CC HSSP; Q60631; IGBR.
CC Genew; HGNC:674; ARHGAP4.
CC MIM; 300023; -
CC GO; GO:0005737; C:cytoplasm; TAS.
CC GO; GO:0005100; F:Rho GTPase activator activity; TAS.
CC GO; GO:0005070; F:SH3/SH2 adaptor protein activity; TAS.
CC GO; GO:0007011; F:regulation of cytoskeleton; TAS.
CC GO; GO:0007266; P:Rho protein signal transduction; TAS.
CC InterPro; IPR001060; Cdc15_Fes_CIP4.
CC InterPro; IPR008936; Rho_GAP.
CC InterPro; IPR00198; RhoGAP.
CC InterPro; IPR01452; SH3.
CC Pfam; PF00611; FCH; 1.
CC Pfam; PF00620; RHOGAP; 1.
CC Pfam; PF00018; SH3; 1.
CC ProDom; PD000066; SH3; 1.
CC SMART; SM00055; FCH; 1.
CC SMART; SM00324; RhoGAP; 1.
CC SMART; SM00326; SH3; 1.
CC PROSITE; PS50133; FCH; 1.
CC PROSITE; PS50238; RHOGAP; 1.
CC PROSITE; PS50002; SH3; 1.
CC GTPase activation; SH3 domain; Coiled coil.
FT DOMAIN 15 88 FCH
FT DOMAIN 128 195 COILED COIL (POTENTIAL).
FT DOMAIN 507 695 RHO-GAP.
FT DOMAIN 746 805 SH3.
FT CONFLICT 609 609 D -> A (IN REF. 2).
FT CONFLICT 731 731 D -> E (IN REF. 2).
SQ SEQUENCE 946 AA; 105055 MM; 8BB291410BCA7DCA CRC64;

Query Match 1.1%; Score 7; DB 1; Length 946;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 498 SLSPPLH 504
DB 95 SLSPPLH 101

RESULT 36
MLB_MYCTU STANDARD; PRT; 966 AA.
ID MLB_MYCTU
AC 053653;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Putative membrane protein mmp11.
GN Mmp11 OR RV0202C OR Wt0212 OR Mtv033.10C OR MB0208C.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
CX NCBI_TaxID=1773; 1765;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,

```

RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Suleson J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RN Nature 393:537-544(1998).
 [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=22206494; PubMed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwin M., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Mayhew L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains.";
 RL J. Bacteriol. 184:5479-5490(2002).
 [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.bovis; STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garner T., Eiglmeyer K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsenpe C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
 RT "The complete genome sequence of Mycobacterium bovis.";
 RT Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: Belongs to the mmpL family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 CC EMBL; AL021928; CA117324.1; -;
 DR EMBL; AB006930; AAK4433.1; -;
 DR EMBL; EX248334; CAD93072.1; -;
 DR PIR; G70838; G70838.
 DR TIGR; MT0212; -;
 DR Tuberculosis; Rv0202c; -;
 DR InterPro: IPR000731; SSD_5TM.
 KW PROSITE; PS50156; SSD; 1.
 KM Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 188 208 POTENTIAL.
 FT TRANSMEM 214 234 POTENTIAL.
 FT TRANSMEM 235 255 POTENTIAL.
 FT TRANSMEM 279 299 POTENTIAL.
 FT TRANSMEM 311 331 POTENTIAL.
 FT TRANSMEM 373 393 POTENTIAL.
 FT TRANSMEM 409 429 POTENTIAL.
 FT TRANSMEM 527 547 POTENTIAL.
 FT TRANSMEM 577 577 POTENTIAL.
 FT TRANSMEM 595 615 POTENTIAL.
 FT TRANSMEM 646 666 POTENTIAL.
 FT TRANSMEM 668 688 POTENTIAL.
 SO SEQUENCE 966 AA; 103502 MW; 3E8BF0327CBEA2DA CRC64;

Query Match 1.1%; Score 7; DB 1; Length 966;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 351 GORPVP 357
 Db 798 GORPVP 804

RESULT 37
 KPI7_MOUSE

ID KPI7_MOUSE STANDARD; PRT; 1038 AA.
 AC 099P8;
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DE 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Kinesin-like protein Kif17 (MmKif17).
 GN KIF17.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RP [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=20307907; PubMed=10846156;
 RA Seton M., Nakagawa T., Seog D.-H., Hirokawa N.;
 RT "Kinesin superfamily motor protein Kif17 and mLin-10 in NMDA
 RT receptor-containing vesicle transport.";
 RL Science 288:1796-1802(2000).
 CC -1- FUNCTION: Transports vesicles containing N-methyl-D-aspartate
 CC (NMDA) receptor 2B along microtubules.
 CC -1- SUBUNIT: Interacts with LIN-10 PDZ domain.
 CC -1- TISSUE SPECIFICITY: Neuronal-specific.
 CC -1- SIMILARITY: Belongs to the kinesin-like protein family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 CC EMBL; AB008667; BAB21099.1; -;
 DR HSSP; P17119; 3KAR.
 DR MGD; MGI:1098229; Kif17.
 DR GO; GO:0005871; C:kinesin complex; IDA.
 DR GO; GO:0003777; F:microtubule motor activity; IDA.
 DR GO; GO:0005515; F:protein binding; IPI.
 DR GO; GO:0007017; F:microtubule-based process; IDA.
 DR GO; GO:0016192; P:vesicle-mediated transport; IDA.
 DR InterPro: IPR001752; kinesin_motor.
 DR Pfam: PF00225; kinesin_1.
 DR PRINTS; PR00380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
 DR PROSITE; PS00677; KINESIN_MOTOR_DOMAIN2; 1.
 KM Motor protein; Microtubule; ATP-binding; Coiled coil; Transport;
 KW Protein transport.
 FT DOMAIN 1 265 KINESIN-MOTOR.
 FT DOMAIN 346 470 COILED COIL (POTENTIAL).
 FT DOMAIN 748 855 COILED COIL (POTENTIAL).
 FT NP BIND 91 98 ATP (POTENTIAL).
 SO SEQUENCE 1038 AA; 116372 MW; 2BED852A3AFBD46 CRC64;

Query Match 1.1%; Score 7; DB 1; Length 1038;
 Best Local Similarity 100.0%; Pred. No. 2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 343 DSATSEA 349
 Db 547 DSATSEA 553

RESULT 38
 UBP7_YEAST
 ID UBP7_YEAST STANDARD; PRT; 1071 AA.
 AC P404E3;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Ubiquitin carboxyl-terminal hydrolase 7 (EC 3.1.2.15) (Ubiquitin
 DE thioesterase 7) (Ubiquitin-specific processing protease 7)

```

DE (Deubiquitinating enzyme 7).
GN UBP7 OR YII156W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RX MEDLINE=97313266; PubMed=9169870;
RA Churcher C.M., Bowman S., Baddock K., Bankier A., Brown D.,
RA Hittingworth T., Connor R., Devlin K., Gentles S., Hamlin N.,
RA Harris D.E., Hortsnell T., Hunt S., Jagels K., Jones M., Lye G.,
RA Moulé S., Odeh C., Pearson D., Rajandream M.A., Rice P., Rowley N.,
RA Skelton J., Smith V., Walsh S., Whitehead S., Barrett B.G.,
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome IX.",
RL Nature 387:84-87(1997).
CC -1- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
CC ubiquitin + a thiol.
CC -1- SIMILARITY: Belongs to peptidase family C19.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z38059; CA86122.1; -.
DR PIR; S48378; S48378.
DR GeneOnline; 139691; -.
DR MEROPS; C19, UPW; -.
DR SGD; S0001418; UBP7.
DR GO; GO:0005737; C-cytoplasm; IC.
DR InterPro; IPR001394; Peptidase_C19.
DR InterPro; IPR001763; Rhodanese-like.
DR Pfam; PF00581; Rhodanese; 1.
DR SMART; SM00450; RHOD; 1.
DR PROSITE; PS00972; UCH_2_1; 1.
DR PROSITE; PS00973; UCH_2_2; 1.
DR PROSITE; PS0235; UCH_2_3; 1.
KW Ubl conjugation pathway; Hydrolase; Thiol protease; Multigene family.
FT ACT SITE 618 BY SIMILARITY.
FT ACT SITE 1006 BY SIMILARITY.
FT ACT SITE 1014 BY SIMILARITY.
FT ACT SITE 1014 BY SIMILARITY.
SQ SEQUENCE 1071 AA; 123133 MW; 82683A01063CEC74 CRC64;
Query Match 1.1%; Score 7; DB 1; Length 1071;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 24 EXVLLID 30
DB 330 EXVLLID 336
RESULT 39
CYA7_MOUSE STANDARD; PRT; 1099 AA.
AC PS1829;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Adenylate cyclase, type VII (EC 4.6.1.1) (ATP pyrophosphate-lyase)
DE (Adenylate cyclase).
GN ADCY7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC TISSUE=Lymphoma;
RX MEDLINE=95050703; PubMed=7961850;
RA Watson P.A., Krupinski J., Kempinski A.M., Frankenfeld C.D.;
RT "Molecular cloning and characterization of the type VII isoform of
RT mammalian adenylate cyclase expressed widely in mouse tissues and in
RT S49 mouse lymphoma cells."
RL J. Biol. Chem. 269:28893-28898(1994).
CC -1- FUNCTION: This is a membrane-bound, calcium-inhibitable adenylate
CC cyclase.
CC -1- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.
CC -1- COFACTOR: Binds 2 magnesium ions per subunit (By similarity).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: Most abundant in heart, spleen and lung.
CC -1- DOMAIN: Composed of two homologous domains.
CC -1- SIMILARITY: Belongs to the adenylate cyclase class-4/guanylate
CC cyclase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U12919; AAAS7554.1; -.
DR PIR; A55405; A55405.
DR HSSP; P26769; IAB8.
DR MGD; MGI:102891; Adcy7.
DR InterPro; IPR001054; G_cyclase.
DR Pfam; PF00211; guanylate_cyc; 2.
DR SMART; SM00044; CYCC; 2.
DR PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.
DR PROSITE; PS00452; GUANYLATE_CYCLASES_2; 2.
KW lysase; GMP biosynthesis; Transmembrane; Glycoprotein; Repeat;
KW Metal-binding; Magnesium.
FT DOMAIN 1 33 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 34 54 POTENTIAL.
FT TRANSMEM 63 83 POTENTIAL.
FT TRANSMEM 95 117 POTENTIAL.
FT TRANSMEM 122 142 POTENTIAL.
FT TRANSMEM 147 167 POTENTIAL.
FT TRANSMEM 178 198 POTENTIAL.
FT DOMAIN 199 595 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 596 616 POTENTIAL.
FT TRANSMEM 621 641 POTENTIAL.
FT TRANSMEM 670 689 POTENTIAL.
FT TRANSMEM 719 738 POTENTIAL.
FT TRANSMEM 747 766 POTENTIAL.
FT TRANSMEM 813 833 POTENTIAL.
FT DOMAIN 834 1099 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 834 1099 POTENTIAL.
FT METAL 286 286 MAGNESIUM 1 AND 2 (BY SIMILARITY).
FT METAL 287 287 MAGNESIUM 2 (VIA CARBOXYL OXYGEN) (BY
FT SIMILARITY).
FT METAL 330 330 MAGNESIUM 1 AND 2 (BY SIMILARITY).
FT CARBOHYD 702 702 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 1099 AA; 122767 MW; ADF3B3559CD11B96 CRC64;
Query Match 1.1%; Score 7; DB 1; Length 1099;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 127 EFSRCFP 133
DB 644 EFSRCFP 650
RESULT 40
LEPR_MACMU STANDARD; PRT; 1163 AA.
AC Q9MYI0; Q9MYK9; Q9MYL1; Q9MYL2;
DT 28-FEB-2003 (Rel. 41, Created)

```

DT 28-FEB-2003 (Rel. 41, last sequence update)
 DT 15-MAR-2004 (Rel. 43, last annotation update)
 DE Leptin receptor precursor (LEP-R) (OB receptor) (OB-R).
 GN LEP-R OR OBR.
 OS Macaca mulatta (Rhesus macaque).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
 OC Cercopithecinae; Macaca.
 NCBI_TaxID=9544;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS A AND B).
 RC TISSUE=Adipose tissue;
 RX MEDLINE=98408931; PubMed=9738551;
 RA Hotta K., Gustafson T.A., Ortmeyer H.K., Bodkin N.L., Hansen B.C.;
 RT "Monkey leptin receptor mRNA: sequence, tissue distribution, and mRNA
 expression in the adipose tissue of normal, hyperinsulinemic, and type
 2 diabetic rhesus monkeys.";
 RT Obes. Res. 6:353-360(1998).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS A AND B).
 RC TISSUE=Adipose tissue;
 RA Hotta K., Gustafson T.A., Ortmeyer H.K., Bodkin N.L., Hansen B.C.;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 RP [3]
 RP SEQUENCE FROM N.A.
 RA Gustafson T.A., Ortmeyer H.K., Bodkin N.L., Hansen B.C.;
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Receptor for Obesity Factor (leptin). Involved in the
 regulation of fat metabolism and in a hematopoietic pathway
 required for normal lymphopoiesis. May play a role in
 reproduction.
 CC -1- FUNCTION: The short form (isoform A) may act to transport leptin
 to the cerebrospinal fluid (By similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=B;
 CC IsoId=Q9MYL0-1; Sequence=Displayed;
 CC Name=A;
 CC IsoId=Q9MYL0-2; Sequence=VSP_001695, VSP_001696;
 CC -1- TISSUE SPECIFICITY: Widely expressed. High expression of isoform B
 in liver, adipose tissue, hypothalamus and choroid plexus.
 CC -1- DOMAIN: The cytoplasmic domain may be essential for intracellular
 signal transduction by activation of JAK tyrosine kinase and
 STATs.
 CC -1- PTM: Phosphorylated on two tyrosine residues. Tyr-984 may be the
 major site of phosphorylation. Phosphorylation on both sites is
 required for full activity (By similarity).
 CC -1- SIMILARITY: Belongs to the type I cytokine family of receptors.
 CC Subfamily 2.
 CC -1- SIMILARITY: Contains 3 fibronectin type III domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AF225874; AAF5388.1; -;
 CC EMBL: AF225875; AAF5389.1; -;
 CC EMBL: AF225873; AAF5387.1; -;
 CC EMBL: AF222960; AAF34683.1; -;
 CC HSSP: P16471; 1BP3.
 DR InterPro: IPR002996; CRIA.
 DR InterPro: IPR008957; FN_III-like.
 DR InterPro: IPR003964; FN_III.
 DR InterPro: IPR003529; Hemtopoptn_L_F2.
 DR InterPro: IPR003531; Hemtopoptn_S_F1.
 DR Pfam: PF00041; fn3; 2.
 DR SMART: SM00060; FN3; 4.
 DR PROSITE: PS01353; HEMATOPO_REC_L_F2; 1.

DR PROSITE; PS01353; HEMATOPO_REC_S_F1; 1.
 KM Obesity; Receptor; Transmembrane; Glycoprotein; Signal; Repeat;
 KW Alternative splicing; Phosphorylation.
 FT SIGNAL 1 21
 FT CHAIN 22 1163
 FT DOMAIN 22 837
 FT TRANSMEM 838 860
 FT DOMAIN 861 1163
 FT DOMAIN 235 318
 FT DOMAIN 535 621
 FT DOMAIN 736 821
 FT DISULFID 37 88
 FT DISULFID 87 97
 FT DISULFID 129 140
 FT DISULFID 184 194
 FT DISULFID 350 410
 FT DISULFID 411 416
 FT DISULFID 471 526
 FT DISULFID 466 496
 FT MOD_RES 984 984
 FT MOD_RES 1139 1139
 FT VARSPPLIC 890 894
 FT VARSPPLIC 895 1163
 FT FT
 FT CONFLICT 889 889
 FT SEQUENCE 1163 AA; 132295 MW; 6B7889108F5B1895 CRC64;
 SQ
 Query Match 1.1%; Score 7; DB 1; Length 1163;
 Best local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 371 LEDSPLY 377
 DB 167 LEDSPLY 173
 RESULT 41
 LEP-R HUMAN STANDARD; PRT; 1165 AA.
 ID P48357; Q13592; Q13593; Q13594; Q92919; Q92920; Q92921;
 AC P48357; Q13592; Q13593; Q13594; Q92919; Q92920; Q92921;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Leptin receptor precursor (LEP-R) (OB receptor) (OB-R) (HUB219).
 GN LEP-R OR OBR OR DB.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS B AND E).
 RC TISSUE=Brain;
 RX MEDLINE=96128129; PubMed=8548812;
 RA Taregetia L.A., Dembeki M., Weng X., Deng N., Chieppier J.,
 RA Devos R., Richards G.-J., Campfield L.A., Clark F.T., Deeds J.,
 RA Muir C., Sanker S., Moriarty A., Moore K.J., Smutko J.S.,
 RA Mays G.G., Wolf E.A., Monroe C.A., Tepper R.I.;
 RT "Identification and expression cloning of a leptin receptor, OB-R,";
 RT Cell 83:1263-1271(1995).
 RL Cell 83:1263-1271(1995).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS A, B AND C).
 RC TISSUE=Fetal liver;
 RX MEDLINE=96398968; PubMed=8805376;
 RA Bennett B.D., Solar G.P., Yuan J.Q., Mathias J., Thomas G.R.,
 RA Matthews W.;
 RT "A role for leptin and its cognate receptor in hematopoiesis,";
 RT Curr. Biol. 6:1170-1180(1996).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM B), AND VARIANTS ARG-109 AND ARG-223.
 RX MEDLINE=97301763; PubMed=9158141;
 RA Thompson D.B., Ravussin E., Bennett P.H., Bogardus C.;

RT "Structure and sequence variation at the human leptin receptor gene in
 RT lean and obese Pima Indians.";
 RL Hum. Mol. Genet. 6:675-679(1997).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM A).
 RX MEDLINE=97215244; PubMed=9061609;
 RA Luo S.-M., Di Marco F., Levin N., Armani M., Xie M.H., Nelson C.,
 RT Bennett G.L., Williams M., Spencer S.A., Gurney A., de Sauvage F.J.;
 RT "Cloning and characterization of a human leptin receptor using a
 RT biologically active leptin immunoadhesin.";
 RL J. Mol. Endocrinol. 18:77-85(1997).
 RN [5]
 RP SEQUENCE FROM N.A. (ISOFORMS A; C AND D).
 RC TISSUE=Fetal liver;
 RX MEDLINE=96206286; PubMed=8616721;
 RA Cioffi J.A., Shafer A.W., Zupancic T.J., Smith-Gbur J., Mikhail A.,
 RT Platika D., Snodgrass H.R.;
 RT "Novel B219/OB receptor isoforms: possible role of leptin in
 RT hematopoiesis and reproduction.";
 RL Nat. Med. 2:585-589(1996).
 RN [6]
 RP ALTERNATIVE SPLICING DUE TO AN ENDOGENOUS RETROVIRUS.
 RX MEDLINE=99128223; PubMed=9923994;
 RA Kapitonov V.V., Uirka J.;
 RT "The long terminal repeat of an endogenous retrovirus induces
 RT alternative splicing and encodes an additional carboxy-terminal
 RT sequence in the human leptin receptor.";
 RL J. Mol. Evol. 48:248-251(1999).
 RN [7]
 RP CARBOHYDRATE-LINKAGE SITES, DISULFIDE BONDS, AND PARTIAL SEQUENCE.
 RX MEDLINE=99003211; PubMed=9786864;
 RA Hanin M., Arakawa T., Bures E.J., Young Y., Hui J.O., Rohde M.F.,
 RT Weicher A.A., Horan T.;
 RT "Human leptin receptor. Determination of disulfide structure and
 RT N-glycosylation sites of the extracellular domain.";
 RL J. Biol. Chem. 273:28691-28699(1998).
 RN [8]
 RP VARIANT ARG-223.
 RX MEDLINE=96270489; PubMed=8666155;
 RA Considine R.V., Considine E.L., Williams C.J., Hyde T.M., Caro J.F.;
 RT "The hypothalamic leptin receptor in humans: identification of
 RT incidental sequence polymorphisms and absence of the db/db mouse and
 RT fa/fa rat mutations.";
 RL Diabetes 45:992-994(1996).
 RN [9]
 RP VARIANTS ARG-109; ARG-204; ARG-223 AND ASN-656.
 RX MEDLINE=97289527; PubMed=9144432;
 RA Echalwid S.M., Soerensen T.D., Soerensen T.I., Tybjæerg-Hansen A.,
 RT Andersen T., Chung W.K., Leibel R.L., Pedersen O.;
 RT "Amino acid variants in the human leptin receptor: lack of association
 RT to juvenile onset obesity.";
 RL Biochem. Biophys. Res. Commun. 233:248-252(1997).
 RN [10]
 RP VARIANTS ARG-109; ARG-223 AND ASN-656.
 RX MEDLINE=97431549; PubMed=9287054;
 RA Chung W.K., Power-Kehoe L., Chua M., Chu F., Aronow L., Huma Z.,
 RT Southern M., Udall J.N., Kahle B., Leibel R.L.;
 RT "Exonic and intronic sequence variation in the human leptin receptor
 RT gene (LEPR).";
 RL Diabetes 46:1509-1511(1997).
 RN [11]
 RP VARIANTS ARG-109; ARG-223; ASN-656 AND THR-675.
 RX MEDLINE=99075638; PubMed=9860295;
 RA Roth H., Korn T., Rosenkranz K., Himney A., Ziegler A., Kunz J.,
 RT Siegfried W., Mayer H., Hebebrand J., Grzeschik K.-H.;
 RT "Transmission disequilibrium and sequence variants at the leptin
 RT receptor gene in extremely obese German children and adolescents.";
 RL Hum. Genet. 103:540-546(1998).
 RN [12]
 RP VARIANTS ARG-109; ARG-223 AND ASN-656.
 RX MEDLINE=97318795; PubMed=9175732;
 RA Gotoda T., Manning B.S., Goldstone A.P., Imrie H., Evans A.L.,
 RT Strosberg A.D., McKeligue P.M., Scott J., Altman T.J.;

RT "Leptin receptor gene variation and obesity: lack of association in a
 RT white British male population.";
 RL Hum. Mol. Genet. 6:869-876(1997).
 RN [1]
 RP FUNCTION: Receptor for obesity factor (leptin). Involved in the
 CC regulation of fat metabolism and in a hematopoietic pathway
 CC required for normal lymphopoiesis. May play a role in
 CC reproduction.
 CC [1]
 CC FUNCTION: The short form (isoform A) may act to transport leptin
 CC to the cerebrospinal fluid (by similarity).
 CC [1]
 CC SUBCELLULAR LOCATION: Type I membrane protein. Except for form E
 CC which could be soluble.
 CC [1]
 CC ALTERNATIVE PRODUCTS:
 CC Name=B; Synonyms=13.2, OB6D;
 CC Name=A; Synonyms=6.4, HUB219.3;
 CC Name=C; Synonyms=12.1, OBRA;
 CC Name=D; Synonyms=HUB219.2;
 CC Name=E; Synonyms=VSP_001693, VSP_001694;
 CC Name=F;
 CC IsoId=P48357-5; Sequence=VSP_001688;
 CC [1]
 CC TISSUE SPECIFICITY: Isoform A expressed in fetal liver and in
 CC hematopoietic tissues and choroid plexus. In adults highest
 CC expression in heart, liver, small intestine, prostate and ovary.
 CC Low level in lung and kidney. Isoform B is highly expressed in
 CC hypothalamus.
 CC [1]
 CC DOMAIN: The cytoplasmic domain may be essential for intracellular
 CC signal transduction by activation of JAK tyrosine kinase and
 CC STATs.
 CC [1]
 CC PM: Phosphorylated on two tyrosine residues. Tyr-966 may be the
 CC major site of phosphorylation. Phosphorylation on both sites is
 CC required for full activity (by similarity).
 CC [1]
 CC SIMILARITY: Belongs to the type I cytokine family of receptors.
 CC Subfamily 2.
 CC [1]
 CC SIMILARITY: Contains 3 fibronectin type III domains.
 CC [1]
 CC SIMILARITY: Contains 1 immunoglobulin-like domain.
 CC [1]
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC [1]
 CC EMBL; U43168; AAA93015.1; -;
 CC EMBL; U66495; AAB07495.1; -;
 CC EMBL; U66496; AAB07496.1; -;
 CC EMBL; U66497; AAB07497.1; -;
 CC EMBL; U59263; AAB09673.1; -;
 CC EMBL; U59248; AAB09673.1; JOINED.
 CC EMBL; U59249; AAB09673.1; JOINED.
 CC EMBL; U59250; AAB09673.1; JOINED.
 CC EMBL; U59252; AAB09673.1; JOINED.
 CC EMBL; U59253; AAB09673.1; JOINED.
 CC EMBL; U59254; AAB09673.1; JOINED.
 CC EMBL; U59255; AAB09673.1; JOINED.
 CC EMBL; U59256; AAB09673.1; JOINED.
 CC EMBL; U59257; AAB09673.1; JOINED.
 CC EMBL; U59258; AAB09673.1; JOINED.
 CC EMBL; U59259; AAB09673.1; JOINED.
 CC EMBL; U59260; AAB09673.1; JOINED.
 CC EMBL; U59261; AAB09673.1; JOINED.
 CC EMBL; U59262; AAB09673.1; JOINED.
 CC EMBL; U50748; AAC23650.1; -;
 CC EMBL; U52912; AAC50505.1; -;
 CC EMBL; U52913; AAC50510.1; -;
 CC EMBL; U52914; AAC50511.1; -;
 CC HSSP; P16471; IBP3.
 CC GeneW; HGNC:6554; LEPR.
 CC MIW; 601007; -;


```
DR GO; GO:0004888; P:transmembrane receptor activity; TAS.
DR GO; GO:0007166; P:cell surface receptor linked signal transdu. . .; TAS.
DR GO; GO:0007275; P:development; TAS.
DR GO; GO:0006112; P:energy reserve metabolism; TAS.
DR InterPro; IPR002996; CRA.
DR InterPro; IPR008957; FN_III-like.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003529; Hemicoptn_L_F2.
DR InterPro; IPR007110; Ig-like.
DR Pfam; PF00041; fn3; 2.
DR SMART; SM00060; FN3; 4.
DR PROSITE; PS01353; HEMATOPO_REC_L_F2; 1.
DR PROSITE; PS50835; IG LIKE; 1.
KM Obesity; Receptor; Transmembrane; Glycoprotein; Signal;
KW Immunoglobulin domain; Repeat; Alternative splicing; Phosphorylation;
KM Polymorphism.
FT SIGNAL 1 21
FT CHAIN 22 1165
FT DOMAIN 22 839
FT TRANSMEM 840 862
FT DOMAIN 863 1165
CYTOPLASMIC (POTENTIAL).

Query Match 1.1%; Score 7; DB 1; Length 1165;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 371 LEDSPLV 377
DB 169 LEDSPLV 175

RESULT 42
DHX8_ARATH STANDARD; PRT; 1168 AA.
ID DHX8; Q9LRV0;
AC Q38953; Q9LRV0;
DT 15-DEC-1998 (Rel. 37, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Putative pre-mRNA splicing factor ATP-dependent RNA helicase.
GN ATG36560 OR MEI16.8.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC Eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20277480; PubMed=10819329;
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT features of the regions of 4,504,864 bp covered by sixty P1 and TAC
RT clones."
RL DNA Res. 7:131-135(2000).
CC -1- FUNCTION: Could act late in the splicing of pre-mRNA and mediate
CC the release of the spliced mRNA from spliceosomes (By similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY. DEAD
CC SUBFAMILY. DDX8/PRP22 ORTHOLOG.
CC -1- SIMILARITY: Contains 1 SI motif domain.
CC -1- CAUTION: Ref.1 sequences differ from that shown due to a
CC frameshift in position 57.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X98130; CAA66825.1; ALT FRAME.
DR EMBL; X97970; CAA66613.1; ALT FRAME.
DR EMBL; AB028611; BAB01838.1; -.
DR HSSP; P05055; ISRO.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR002464; DEAD box.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR007502; Helicase_dom.
DR InterPro; IPR008984; Nucleic_acid_OB.
DR InterPro; IPR003029; S1.
DR Pfam; PF04408; HA2; 1.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00575; S1; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELIC; 1.
DR SMART; SM00316; S1; 1.
DR PROSITE; PS00690; DEAD_ATP_HELICASE; 1.
DR PROSITE; PS50126; S1; 1.
KM Hypochemical protein; mRNA processing; mRNA splicing; Helicase;
KW ATP-binding; Nuclear protein.
FT DOMAIN 214 283
FT NE_BIND 538 545
FT SITE 635 638
FT FT 777 780
FT DOMAIN POLY-PRO.
SQ SEQUENCE 1168 AA; 134156 MW; B3632DEA4A7A7690C CRC64;
```

```
Query Match 1.1%; Score 7; DB 1; Length 1168;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 KMKRRL 58
DB 653 KMKRRL 659

RESULT 43
AIP1_MOUSE STANDARD; PRT; 1275 AA.
ID AIP1_MOUSE
AC Q9W0L1; Q8BYT1; Q8CAB5;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Atrophin-1 interacting protein 1 (Membrane associated guanylate kinase
DE inverted-2) (MAGI-2) (Activin receptor interacting protein 1)
DE (Acvrip1).
GN AIP1 OR MAGI2 OR AIP1 OR ACVRINP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2). INTERACTION WITH MADH2, MADH3
RP AND ACV22, AND IDENTIFICATION IN A COMPLEX WITH ACV22, ACV1B AND
RP MADH3.
RC STRAIN=ICR; TISSUE=Brain;
RX MEDLINE=20148748; PubMed=10681527;
RA Shoji H., Tsuchida K., Kishi H., Yamakawa N., Matsuzaki T., Liu Z.,
RA Nakamura T., Sugino H.;
RT "Identification and characterization of a PDZ protein that interacts
RT with activin types II receptors."
RL J. Biol. Chem. 275:5485-5492(2000).
RN [2]
RP SEQUENCE OF 159-1112 FROM N.A. (ISOFORM 3).
RC STRAIN=57BL/6J; TISSUE=Hypothalamus, and spinal cord;
RX MEDLINE=22354683; PubMed=12466851;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaido I., Oosato N., Salto R., Suzuki H., Yamakawa I., Kiyosawa H.,
```

RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojohori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schiraldi L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bratt D., Bruscia V., Chochua C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gutlich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Komagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Meglert D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagatsuma T., Numa K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Sempole C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Varvaro R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilmink L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Kono H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterson R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.,
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs";
RL Nature 420:563-573(2002).
CC -1- FUNCTION: Seems to act as scaffold molecule at synaptic junctions
CC by assembling neurotransmitter receptors and cell adhesion
CC proteins. May play a role in regulating activin-mediated signaling
CC in neuronal cells. Enhances the ability of p15N to suppress AKT1
CC activation (By similarity).
CC -1- SUBUNIT: Interacts via its second PDZ domain with PTEN unphosphorylated C-
CC terminus (By similarity). Interacts through its guanylate kinase
CC domain with DGAP1 (By similarity). Interacts through the PDZ
CC domains with GRIN2A, GRIN2B and NLGN1 (By similarity). Interacts
CC with CTNND2, CTNND1 and MAGI1-1 (By similarity). Interacts with
CC ACVR2, MADH2 and MADH3. Part of a complex consisting of Alpi,
CC ACVR2, ACVR1B and MADH3.
CC -1- SUBCELLULAR LOCATION: Membrane-associated in synaptosomes (By
CC similarity).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=1; Synonyms=long;
CC IsoId=Q9WVQ1-1; Sequence=Displayed;
CC Name=2; Synonyms=short;
CC IsoId=Q9WVQ1-2; Sequence=VSP_008436;
CC Note=Major;
CC Name=3;
CC IsoId=Q9WVQ1-3; Sequence=VSP_008437;
CC Note=No experimental confirmation available;
CC -1- TISSUE SPECIFICITY: Specifically expressed in brain.
CC -1- SIMILARITY: Belongs to the MAGUK family.
CC -1- SIMILARITY: Contains 1 guanylate kinase-like domain.
CC -1- SIMILARITY: Contains 6 PDZ/DHR domains.
CC -1- CAUTION: Ref.2 (BAC29987) sequence differs from that shown due to
CC a frameshift in position 102.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
CC or send an email to license@isb-sdb.ch).
CC -----
CC EMBL; AB029485; BAA82294.1; -;
CC EMBL; AK039336; BAC30321.1; ALT_INT.
CC EMBL; AK038407; BAC29987.1; ALT_FRAME.
CC PIR; P10546; P10698.
CC HSP; P29476; IQAV.

DR MGD: MG1.1354953; Acyrinpl.
DR GO; GO:0004871; F-signal transducer activity; IPI.
DR GO; GO:0007165; P-signal transduction; IPI.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001202; WW_Rsp5_WMP.
DR Pfam; PF00595; PDZ; 5.
DR Pfam; PF00397; WW; 2.
DR SMART; SM00228; PDZ; 5.
DR SMART; SM00456; WW; 2.
DR PROSITE; PS00856; GUANYLATE KINASE 1; 1.
DR PROSITE; PS50052; GUANYLATE KINASE 2; 1.
DR PROSITE; PS50106; PDZ; 6.
DR PROSITE; PS01159; WW_DOMAIN 1; 2.
DR PROSITE; PS50020; WW_DOMAIN 2; 2.
KW Repeat, Alternative splicing.
FT DOMAIN 17 101 PDZ 1.
FT DOMAIN 109 285 GUANYLATE KINASE.
FT DOMAIN 301 334 WW 1.
FT DOMAIN 347 380 WW 2.
FT DOMAIN 425 509 PDZ 2.
FT DOMAIN 604 682 PDZ 3.
FT DOMAIN 777 859 PDZ 4.
FT DOMAIN 919 1009 PDZ 5.
FT DOMAIN 1139 1221 PDZ 6.
FT VARSPPLIC 1 163 Missing (in isoform 2).
FT VARSPPLIC 1229 1275 /FtId=VSP_008436.
FT VARSPPLIC 1275 1275 WPSLSLSCMSDGHGSFYFLGHPKDTNPFGVLEPLP
FT POACRK -> AFHGFHLCSAFSVF (in isoform 3).
FT /FtId=VSP_008437.
SQ SEQUENCE 1275 AA; 140918 MW; F17DC552517806354 CRC64;
Query Match 1.1%; Score 7; DB 1; Length 1275;
Beet Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 17 ALBESGT 23
DB 171 ALBESGT 177
RESULT 44
AIP1_RAT STANDARD; PRT; 1277 AA.
ID AIP1_RAT
AC 088382; Q9R271;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Atrophin-1 interacting protein 1 (Membrane associated guanylate kinase
DE inverted-2) (MAGI-2) (Synaptic scaffolding molecule) (S-SCAM).
GN AIP1 OR MAGI2 OR SSCAM.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1), FUNCTION, SUBCELLULAR LOCATION, TISSUE
RP SPECIFICITY, AND INTERACTION WITH DGAP1, NLGN1 AND GRIN2A.
RX MEDLINE=96361985; PubMed=9694864;
RA Hirao K., Hata Y., Ide N., Takeuchi M., Irie M., Yao I., Deguchi M.,
RA Toyoda A., Sudhof T.C., Takai Y.,
RT "A novel multiple PDZ domain-containing molecule interacting with
RT N-methyl-D-aspartate receptors and neuronal cell adhesion proteins".
RL J. Biol. Chem. 273:21105-21110(1998).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND ALTERNATIVE SPLICING (ISOFORM 3).
RX MEDLINE=20112872; PubMed=10644767;
RA Hirao K., Hata Y., Yao I., Deguchi M., Kawabe H., Mizoguchi A.,
RA Takai Y.,
RT "Three isoforms of synaptic scaffolding molecule and their
RT characterization: Multimerization between the isoforms and their
RT interaction with N-methyl-D-aspartate receptors and SAP90/PSD-95-
RT associated protein.";
RL J. Biol. Chem. 275:2966-2972(2000).

RN [3]
 RP INTERACTION WITH MAGUI-1.
 RX MEDLINE=9922314; PubMed=10207009;
 RA Yao I., Hata Y., Ide N., Hiroo K., Deguchi M., Nishiohka H.,
 RA Mizoguchi A., Takai Y.,
 RT "MAGUI-1, a novel neuronal membrane-associated guanylate
 RT kinase-interacting protein."
 RL J. Biol. Chem. 274:11889-11896(1999).
 RN [4]
 RP INTERACTION WITH CTNNB2.
 RX MEDLINE=99182311; PubMed=10080919;
 RA Ide N., Hata Y., Deguchi M., Hiroo K., Yao I., Takai Y.,
 RT "Interaction of S-SCAM with neural plakophilin-related
 RT Armadillo-repeat protein/delta-catenin."
 RL Biochem. Biophys. Res. Commun. 256:456-461(1999).
 RN [5]
 RP INTERACTION WITH GRID2.
 RX MEDLINE=22477859; PubMed=12589829;
 RA Yip C.C., Muto Y., Kishida H., Hashikawa T., Yano R.,
 RT "PKC regulates the delta2 glutamate receptor interaction with
 RT S-SCAM/MAGI-2 protein."
 RL Biochem. Biophys. Res. Commun. 301:1122-1128(2003).
 CC -1- FUNCTION: Seems to act as scaffold molecule at synaptic junctions
 CC by assembling neurotransmitter receptors and cell adhesion
 CC proteins. May play a role in regulating activin-mediated signaling
 CC in neuronal cells. Enhances the ability of PTEN to suppress AKT1
 CC activation (By similarity).
 CC -1- SUBUNIT: Interacts through guanylate kinase domain its with
 CC DLGAP1. Interacts through the PDZ domains with GRIIN2A, GRID2 and
 CC NG2N1. Interacts with CTNNB2 and MAGUI-1. Interacts with CTNNB1,
 CC ACVR2, MADH2 and MADH3 (By similarity). Part of a complex
 CC consisting of AIP1, ACVR2, ACVR1B and MADH3 (By similarity).
 CC Interacts via its WW domains with DRPLA (By similarity). Interacts
 CC via its second PDZ domain with PTEN unphosphorylated C-terminus
 CC (By similarity).
 CC -1- SUBCELLULAR LOCATION: Membrane-associated in synapses.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Name=1; Synonyms=alpha;
 CC IsoId=088382-1; Sequence=VSP_008438;
 CC Name=2; Synonyms=beta;
 CC IsoId=088382-2; Sequence=VSP_008438;
 CC Name=3; Synonyms=gamma;
 CC IsoId=088382-3; Sequence=VSP_008439;
 CC -1- TISSUE SPECIFICITY: Exclusively expressed in brain.
 CC -1- SIMILARITY: Belongs to the MAGUK family.
 CC -1- SIMILARITY: Contains 1 guanylate kinase-like domain.
 CC -1- SIMILARITY: Contains 6 PDZ/DHR domains.
 CC -1- SIMILARITY: Contains 2 WW domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
 CC or send an email to license@ebi.ac.uk).
 CC -----
 DR EMBL; AF034863; AAC31124.1; -
 DR EMBL; AF130819; AAD31015.1; -
 DR HSSP; P29476; IOAV.
 DR InterPro; IPR008144; Guanylate_kin.
 DR InterPro; IPR008145; Guanylate_kin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001202; WW_Reps_WMP.
 DR Pfam; PF00625; Guanylate_kin; 1.
 DR Pfam; PF00595; PDZ; 6.
 DR Pfam; PF00397; WW; 2.
 DR SMART; SM00072; GUKc; 1.
 DR SMART; SM00228; PDZ; 6.
 DR SMART; SM00456; WW; 2.
 DR SMART; SM00856; GUANYLATE_KINASE_1; 1.
 DR PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
 DR PROSITE; PS50052; GUANYLATE_KINASE_2; 1.

DR PROSITE; PS50106; PDZ; 6.
 DR PROSITE; PS01159; WW DOMAIN 1; 2.
 DR PROSITE; PS50020; WW DOMAIN 2; 2.
 KW Repeat; Alternative splicing.
 FT DOMAIN 17 101 PDZ.
 FT DOMAIN 109 285 GUANYLATE KINASE.
 FT DOMAIN 302 335 WW 1.
 FT DOMAIN 348 381 WW 2.
 FT DOMAIN 426 510 PDZ 1.
 FT DOMAIN 605 683 PDZ 2.
 FT DOMAIN 778 860 PDZ 3.
 FT DOMAIN 920 1010 PDZ 4.
 FT DOMAIN 1141 1223 PDZ 5.
 FT VARSPIC 1 163 Missing (in isoform 2).
 FT VARSPIC 1 223 Missing (in isoform 3).
 FT VARSPIC 1 223 Missing (in isoform 3).
 FT CONFLICT 645 645 /FTId=VSP_008439.
 FT SEQUENCE 1277 AA; 141071 MW; E1A435FF3549DF9 CRC64;
 Query Match 1.1%; Score 7; DB 1; Length 1277;
 Best Local Similarity 100.0%; Pred. No. 2.4e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 17 ALLESGT 23
 Db 171 ALLESGT 177
 RESULT 45
 ID VAR2_SCHPO STANDARD; PRT; 1318 AA.
 AC 009804;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein C2G11.02 in chromosome I.
 GN SPAC2G11.02.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetaceae; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9712;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayles J., Baker S., Baaham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagals K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skellon J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Gymnopoulos B.,
 RA Welljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
 RA Giffenau A., Cadieu E., Dreano S., Gloux S., Lelubre V., Mottier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet L., Gallard C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Sipakovski G.V., Ussery D., Barrett B.G., Nurse P.,
 RT "The genome sequence of Schizosaccharomyces pombe".
 RL Nature 415:871-880(2002).

```

CC -1- SIMILARITY: SOME, TO YEAST YJR041C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z53354; CA91167.1; -.
DR PIR; T38568; T38568.
DR GeneDB; Spombe; SPAC2G11.02; -.
KW Hypothetical protein.
SQ SEQUENCE 1318 AA; 151569 MW; 473556045B2E03A CRC64;

Query Match 1.1%; Score 7; DB 1; Length 1318;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 634 ESTIMSEN 640
Db 469 ESTIMSEN 475

RESULT 46
YSYL_CABEL STANDARD; PRT; 1332 AA.
ID YSYL_CABEL
AC Q20255; Q23468;
DT 01-NOV-1997 (Rel. 35, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein ZK328.7 in chromosome III.
GN ZK328.7/F40H6.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematozoa; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Du Z., Favello A.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
RA [2]
RP REVISIONS.
RA Waterston R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 16 TPR repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U50193; AAA91253.2; -.
DR WormBep; ZK328.7; CE29638.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF00515; TPR; 7.
DR SMART; SM00028; TPR; 9.
KW Hypothetical protein; Repeat; TPR repeat.
FT REPEAT 58
FT REPEAT 91
FT REPEAT 1.
FT REPEAT 2.
FT REPEAT 446
FT REPEAT 582
FT REPEAT 615
FT REPEAT 669
FT REPEAT 702
FT REPEAT 737
FT REPEAT 770
FT REPEAT 772
FT REPEAT 804
FT REPEAT 806
FT REPEAT 837
FT REPEAT 847
FT REPEAT 880
FT REPEAT 894
FT REPEAT 927
FT TPR 9.

```

```

FT REPEAT 929
FT REPEAT 961
FT REPEAT 963
FT REPEAT 985
FT REPEAT 997
FT REPEAT 1029
FT REPEAT 1033
FT REPEAT 1066
FT REPEAT 1213
FT REPEAT 1246
FT REPEAT 1248
FT REPEAT 1280
FT REPEAT 1282
FT REPEAT 1315
SQ SEQUENCE 1332 AA; 152188 MW; 72B78AC864DBE0 CRC64;

Query Match 1.1%; Score 7; DB 1; Length 1332;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 230 IEKAKAS 236
Db 237 IEKAKAS 243

RESULT 47
TCOF_HUMAN STANDARD; PRT; 1411 AA.
ID TCOF_HUMAN
AC Q13428; Q99408; Q99860;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Treacle protein (Treacher Collins syndrome protein).
GN TCOF1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96154183; PubMed=8563749;
RA Dixon J., Edwards S.J., Gladwin A.J., Dixon M.J., Loftus S.R.,
RA Bonner C.A., Koprivnikar K., Wasmuth J.J.;
RT "Positional cloning of a gene involved in the pathogenesis of
RT Treacher Collins syndrome.";
RL Nat. Genet. 12:130-136(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97228900; PubMed=9074926;
RA Dixon J., Edwards S.J., Anderson I., Brass A., Scambler P.J.,
RA Dixon M.J.;
RT "Identification of the complete coding sequence and genomic
RT organization of the Treacher Collins syndrome gene.";
RL Genome Res. 7:223-234(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97250498; PubMed=9096354;
RA Wise C.A., Chiang L.C., Paznekas W.A., Sharma M., Musy M.M.,
RA Ashley J.A., Lovett W., Jabs S.W.;
RT "TCOF1 gene encodes a putative nucleolar phosphoprotein that exhibits
RT mutations in Treacher Collins syndrome throughout its coding
RT region.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:3110-3115(1997).
RN [4]
RP VARIANTS LEU-439; VAL-810; VAL-1313 AND GLY-1355, AND VARIANT TCS
RP ARG-53.
RX MEDLINE=97195537; PubMed=9042910;
RA Edwards S.J., Gladwin A.J., Dixon M.J.;
RT "Type mutational spectrum in Treacher Collins syndrome reveals a
RT predominance of mutations that create a premature-termination
RT codon.";
RL Am. J. Hum. Genet. 60:515-524(1997).
CC -1- DISEASE: Defects in TCOF1 are the cause of Treacher Collins
CC syndrome (TCS) [MIM:154500]. TCS is an autosomal dominant disorder
CC of craniofacial development that occurs with an incidence of
CC 1/50,000 live births. The clinical features of TCS are bilaterally
CC symmetrical and include: (1) abnormalities of the external ears,
CC atresia of the external ear canals, and malformation of the middle
CC ear ossicles, which may result in conductive hearing loss; (2)
CC lateral downward sloping of palpebral fissures, frequently with

```

CC colobomas of the lower eyelids; (3) hypoplasia of the mandible and
 CC zygomatic complex; (4) cleft palate.
 CC - - SIMILARITY: Contains 1 LISH domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sib.ch/announce/>
 CC or send an email to license@1sb-sib.ch).
 CC -----
 DR EMBL; U40847; AAC50903.1; -
 DR EMBL; U76366; AAC51181.1; -
 DR EMBL; U84664; AAC51185.1; -
 DR EMBL; U84640; AAC51185.1; JOINED.
 DR EMBL; U84641; AAC51185.1; JOINED.
 DR EMBL; U84642; AAC51185.1; JOINED.
 DR EMBL; U84643; AAC51185.1; JOINED.
 DR EMBL; U84644; AAC51185.1; JOINED.
 DR EMBL; U84645; AAC51185.1; JOINED.
 DR EMBL; U84646; AAC51185.1; JOINED.
 DR EMBL; U84647; AAC51185.1; JOINED.
 DR EMBL; U84648; AAC51185.1; JOINED.
 DR EMBL; U84649; AAC51185.1; JOINED.
 DR EMBL; U84650; AAC51185.1; JOINED.
 DR EMBL; U84651; AAC51185.1; JOINED.
 DR EMBL; U84652; AAC51185.1; JOINED.
 DR EMBL; U84653; AAC51185.1; JOINED.
 DR EMBL; U84654; AAC51185.1; JOINED.
 DR EMBL; U84655; AAC51185.1; JOINED.
 DR EMBL; U84656; AAC51185.1; JOINED.
 DR EMBL; U84657; AAC51185.1; JOINED.
 DR EMBL; U84658; AAC51185.1; JOINED.
 DR EMBL; U84659; AAC51185.1; JOINED.
 DR EMBL; U84660; AAC51185.1; JOINED.
 DR EMBL; U84661; AAC51185.1; JOINED.
 DR EMBL; U84662; AAC51185.1; JOINED.
 DR EMBL; U84663; AAC51185.1; JOINED.
 DR EMBL; U79659; AAB40722.1; -
 DR EMBL; U79645; AAB40722.1; JOINED.
 DR EMBL; U79646; AAB40722.1; JOINED.
 DR EMBL; U79647; AAB40722.1; JOINED.
 DR EMBL; U79648; AAB40722.1; JOINED.
 DR EMBL; U79649; AAB40722.1; JOINED.
 DR EMBL; U79650; AAB40722.1; JOINED.
 DR EMBL; U79651; AAB40722.1; JOINED.
 DR EMBL; U79652; AAB40722.1; JOINED.
 DR EMBL; U79653; AAB40722.1; JOINED.
 DR EMBL; U79654; AAB40722.1; JOINED.
 DR EMBL; U79655; AAB40722.1; JOINED.
 DR EMBL; U79656; AAB40722.1; JOINED.
 DR EMBL; U79657; AAB40722.1; JOINED.
 DR EMBL; U79658; AAB40722.1; JOINED.
 DR Genew; HGNC:11654; TCOR1.
 DR MIM; 606847; -
 DR MIM; 134500; -
 DR GO; GO:0005730; C:nucleolus; TAS.
 DR GO; GO:0005215; P:transporter activity; TAS.
 DR GO; GO:0001501; P:skeletal development; TAS.
 DR InterPro; IPR006594; LISH.
 DR InterPro; IPR003993; trectacle.
 DR Pfam; PF03946; trectacle.3.
 DR PRINTS; PRO1503; TREACLE.
 DR SMART; SM00667; LISH; 1.
 DR PROSITE; PS50896; LISH; 1.
 KM Disease mutation; Polymorphism.
 FT DOMAIN 6 38 LISH.
 FT DOMAIN 89 97 POLY-GU.
 FT DOMAIN 204 207 POLY-SER.
 FT DOMAIN 616 619 POLY-SER.
 FT DOMAIN 919 924 POLY-SER.
 FT DOMAIN 1285 1289 POLY-LYS.

FT DOMAIN 1375 1386 POLY-LYS.
 FT DOMAIN 1398 1405 POLY-LYS.
 FT VARIANT 53 53 W -> R (in TCS).
 FT VARIANT 439 439 /FTId=VAR_005630.
 FT VARIANT 439 439 P -> L.
 FT VARIANT 810 810 /FTId=VAR_005631.
 FT VARIANT 810 810 A -> V.
 FT VARIANT 1313 1313 /FTId=VAR_005632.
 FT VARIANT 1313 1313 A -> V (in dbSNP:15251).
 FT VARIANT 1355 1355 /FTId=VAR_005633.
 FT VARIANT 1355 1355 D -> G.
 FT CONFLICT 1312 1312 /FTId=VAR_005634.
 FT SEQUENCE 1411 AA; 144312 MM; 3880203D985C2699 CRC64;
 SQ
 Query Match. 1.1%; Score 7; DB 1; Length 1411;
 Best Local Similarity 100.0%; Pred. No. 2.6e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 357 PASVPSV 363
 Db 767 PASVPSV 773
 RESULT 48
 ID A1P1_HUMAN STANDARD; PRT; 1455 AA.
 AC Q86UL8; O60434; O60510; Q86U17; Q9UDQ5; Q9UDU1;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Atrophin-1 interacting protein 1 (Atrophin-1 interacting protein A)
 DE (Membrane associated guanylate kinase inverted-2) (MAGI-2).
 OS Homo sapiens (Human).
 GN A1P1 OR MAGI2 OR KIAA0705.
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1), TISSUE SPECIFICITY, AND INTERACTION
 RP WITH DRPLA.
 RC TISSUE=Brain;
 RX MEDLINE=98333405; PubMed=9647693;
 RA Wood J.D., Yuan J., Margolis R.L., Colomer V., Duan X., Kushi J.,
 RA Kaminsky Z., Kleiderlein J.U., Sharp A.H., Rose C.A.;
 RT "Atrophin-1, the DRPLA gene product, interacts with two families of W
 RT domain-containing proteins.";
 RL Mol. Cell. Neurosci. 11:149-160(1998).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 2).
 RC TISSUE=Brain;
 RX MEDLINE=98403880; PubMed=9734811;
 RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
 RA Kotani H., Nomura N., Ohara O.;
 RA "Prediction of the coding sequences of unidentified human genes. X.
 RA The complete sequences of 100 new cDNA clones from brain which can
 RA code for large proteins in vitro.";
 RL DNA Res. 5:169-176(1998).
 RN [3]
 RP SEQUENCE OF 1-100; 141-179; 350-682; 684-1175 AND 1237-1455 FROM N.A.
 RA Courtney L., Harrison M., Lemmox S., Bourne S., Ozeresky P., Carter T.,
 RA Tin-Mollam A., Hawkins W., Kepler D., Sulston J.E., Eldred J.,
 RA Kozlowicz A., Bemis G., Langston Y., Mead K., Haekenson W., Meyer R.,
 RA Szilh K., Waterston R.;
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP FUNCTION, SUBCELLULAR LOCATION, AND INTERACTION WITH PTEN.
 RX MEDLINE=20226087; PubMed=10760291;
 RA Wu X., Hepler K., Casetelino-Prabhu S., Do D., Kaye M.B., Yuan X.-J.,
 RA Wood J., Ross C., Sawyers C.L., Wang Y.E.;
 RT "Evidence for regulation of the PTEN tumor suppressor by a
 RT membrane-localized multi-PDZ domain containing scaffold protein
 RT MAGI-2.";

Proc. Natl. Acad. Sci. U.S.A. 97:4233-4238(2000).
 [5]
 INTERACTION WITH PTEN.
 RX MEDLINE=21623681; PubMed=11707428;
 RA Valquez F., Grossman S.R., Takahashi Y., Rokaas M.V., Nakamura N.,
 RA Sellers W.R.;
 RT "Phosphorylation of the PTEN tail acts as an inhibitory switch by
 preventing its recruitment into a protein complex.";
 RL J. Biol. Chem. 276:48627-48630(2001).
 CC -1- FUNCTION: Seems to act as scaffold molecule at synaptic junctions
 by assembling neurotransmitter receptors and cell adhesion
 proteins. May play a role in regulating activin-mediated signaling
 in neuronal cells. Enhances the ability of PTEN to suppress AKT1
 activation.
 CC -1- SUBUNIT: Interacts via its WW domains with DRP1A. Interacts via
 its second PDZ domain with PTEN unphosphorylated C-terminus.
 CC Interacts through its guanylate kinase domain with DLGAP1 (By
 similarity). Interacts through the PDZ domain with GRIN2A, GRIN2
 and NLGN1 (By similarity). Interacts with CTNND2, CTNNB1,
 MAGUIIN-1, ACVR2, MADH2 and MADH3 (By similarity). Part of a
 complex consisting of AIP1, ACVR2, ACVR1B and MADH3 (By
 similarity).
 CC -1- SUBCELLULAR LOCATION: Membrane-associated in synaptosomes (By
 similarity).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q86UL8-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q86UL8-2; Sequence=VSP_008435;
 CC Note=No experimental confirmation available;
 CC -1- TISSUE SPECIFICITY: Specifically expressed in brain.
 CC -1- SIMILARITY: Belongs to the MAGUK family.
 CC -1- SIMILARITY: Contains 1 guanylate kinase-like domain.
 CC -1- SIMILARITY: Contains 6 PDZ/DHR domains.
 CC -1- SIMILARITY: Contains 2 WW domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL; AF038563; AAC05370.1; -;
 DR EMBL; AB014605; BAA31680.2; ALT_INIT.
 DR EMBL; AC004808; AAC23438.1; -;
 DR EMBL; AC004945; AAC61488.1; -;
 DR EMBL; AC004990; AAC79151.1; -;
 DR EMBL; AC005246; AAC25530.1; -;
 DR EMBL; AC006043; AAD15413.2; -;
 DR EMBL; AC073200; AAP22360.1; -;
 DR EMBL; AC007237; AAP21886.1; -;
 DR HSSP; P29476; IQAV.
 DR MIM; 606382; -;
 DR InterPro; IPR01202; WW_Rsp5_WWP.
 DR Pfam; PF00625; Guanylate_kin; 1.
 DR Pfam; PF00397; WW; 2.
 DR SMART; SM00456; WW; 1.
 DR PROSITE; PS00856; GUANYLATE_KINASE_1; 1.
 DR PROSITE; PS50052; GUANYLATE_KINASE_2; 1.
 DR PROSITE; PS50106; PDZ; 6.
 DR PROSITE; PS01159; WW_DOMAIN_1; 2.
 DR PROSITE; PS50020; WW_DOMAIN_2; 2.
 KW Repeat; Alternative splicing.
 FT DOMAIN 17 101 PDZ 1.
 FT DOMAIN 109 285 GUANYLATE_KINASE.
 FT DOMAIN 302 335 WW 1.
 FT DOMAIN 348 381 WW 2.
 FT DOMAIN 426 510 PDZ 2.
 FT DOMAIN 605 683 PDZ 3.

FT	DOMAIN	778	860	PDZ 4.
FT	DOMAIN	920	1010	PDZ 5.
FT	DOMAIN	1015	1118	PRO-RICH.
FT	DOMAIN	1147	1229	PDZ 6.
FT	DOMAIN	1340	1430	ALA-RICH.
FT	VARSPLIC	757	771	OOVPRTPSPRMSSG -> R (in isoform 2).
FT	CONFLICT	1234	1234	/FTId=VSP_008435.
FT	CONFLICT	1250	1250	Q -> E (IN REF. 2).
FT	CONFLICT	1291	1291	G -> C (IN REF. 1).
FT	CONFLICT	1383	1383	E -> K (IN REF. 1).
FT	CONFLICT	1389	1394	P -> L (IN REF. 1).
FT	CONFLICT	1401	1401	FAGPGG -> SADPAD (IN REF. 1).
FT	CONFLICT	1411	1411	E -> A (IN REF. 1).
FT	CONFLICT	1414	1415	G -> A (IN REF. 1).
FT	CONFLICT	1420	1420	PG -> SV (IN REF. 1).
FT	CONFLICT	1423	1423	G -> A (IN REF. 1).
FT	CONFLICT	1426	1426	P -> R (IN REF. 1).
FT	CONFLICT	1429	1429	K -> R (IN REF. 1).
FT	CONFLICT	1437	1437	V -> G (IN REF. 1).
FT	CONFLICT	1437	1437	P -> R (IN REF. 1).
FT	SEQUENCE	1455 AA;	158752 MW;	3AF8A14D69738281 CRC64;

Query Match 1.1%; Score 7; DB 1; Length 1455;
 Best Local Similarity 100.0%; Pred. No. 2,7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 ALLBSGT 23
 DB 171 ALLBSGT 177

RESULT 49
 MYHB_HUMAN
 ID MYHB_HUMAN STANDARD; PRT; 1972 AA.
 AC P35749; O00396; O94944; P78422;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Myosin heavy chain, smooth muscle isoform (SMHHC).
 GN MYH1 OR KIA00866.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99425270; PubMed=10493829;
 RA Loftus B.J., Kim U.-J., Sneddon V.P., Kalush P., Brandon R.,
 RA Fuhrmann J., Mason T., Crosby M.L., Barnstead M., Cronin L.,
 RA Deslattes Mays A., Cao Y., Xu R.X., Kang H.-L., Mitchell S.,
 RA Eichler E.E., Harris P.C., Venter J.C., Adams W.D.;
 RT "Genome duplications and other features in 12 Mb of DNA sequence from
 human chromosome 16p and 16q.";
 RL Genomics 60:295-308(1999).
 RN [2]
 RP SEQUENCE FROM N.A., AND REVISIONS TO 1263-1266.
 RC TISSUE=Brain;
 RA Nagase T., Kikuno R., Yamakawa H., Ohara O.;
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 1-1266 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=99156230; PubMed=10048485;
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosewa M.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XII.
 RT The complete sequences of 100 new cDNA clones from brain which code
 for large proteins in vitro.";
 RL Nucleic Acids Res. 5:355-364(1998).
 RN [4]
 RP SEQUENCE OF 885-1972 FROM N.A.
 RX MEDLINE=93263189; PubMed=7684189;
 RA Matsunaka R., Yoshida M.C., Furutani Y., Imamura S., Kanda N.,

RA Yanagisawa M., Maesaki T., Takao A.;
 RT "Human smooth muscle myosin heavy chain gene mapped to chromosomal
 region 16q12.";
 RL Am. J. Med. Genet. 46:61-67(1993).
 RN [5]
 RP SEQUENCE OF 1093-1972 FROM N.A.
 RC TISSUE=Hipocampus;
 RA Okajima K.;
 RL Submitted (NOV-1992) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Muscle contraction.
 CC -1- SUBUNIT: Muscle myosin is a hexameric protein that consists of 2
 CC heavy chain subunits (MHC), 2 alkali light chain subunits (MLC)
 CC and 2 regulatory light chain subunits (MLC-2).
 CC -1- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
 CC -1- TISSUE SPECIFICITY: Smooth muscle; expressed in the umbilical
 CC artery, bladder, esophagus and trachea.
 CC -1- DOMAIN: The rodlike tail sequence is highly repetitive, showing
 CC cycles of a 28-residue repeat pattern composed of 4 heptapeptides,
 CC characteristic for alpha-helical coiled coils.
 CC -1- DISEASE: A chromosomal rearrangement, known as pericentric
 CC inversion inv(16)(p13;q22), produces a fusion protein that
 CC consists of the 165 N-terminal residues of CBF-beta (PEP2) with
 CC the tail region of MYH11. This rearrangement is associated with
 CC acute myeloid leukemia of M4EO subtype.
 CC -1- MISCELLANEOUS: Each myosin heavy chain can be split into 1 light
 CC meromyosin (LMW) and 1 heavy meromyosin (HMM). It can later be
 CC split further into 2 globular subfragments (S1) and 1 rod-shaped
 CC subfragment (S2).
 CC -1- SIMILARITY: Contains 1 myosin-like globular head domain.
 CC -1- SIMILARITY: Contains 1 IQ domain.
 CC -1- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
 CC WWW=http://www.infobio.gen.fr/services/chromancer/genes/MYH11ID43.html".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF001548; AAC31665.1; -;
 CC EMBL; U91323; AAC35212.1; -;
 CC EMBL; AB020673; BA474899.2; ALT. INT.
 CC EMBL; D10667; -; NOT_ANNOTATED_CDS.
 CC EMBL; X69292; CAA49154.1; -;
 CC HSSP; P10587; IBR2.
 CC DR Genew; HGNC:7569; MYH11.
 CC MIM; 160745; -;
 CC DR GO; GO:0005859; C:muscle myosin; TAS.
 CC DR InterPro; IPR000048; IQ_region.
 CC DR InterPro; IPR001609; Myosin_head.
 CC DR InterPro; IPR004009; Myosin_N.
 CC DR InterPro; IPR002928; Myosin_tail.
 CC DR InterPro; IPR002017; Spectrin.
 CC DR Pfam; PF00612; IQ; 1.
 CC DR Pfam; PF00063; myosin_head; 1.
 CC DR Pfam; PF02736; Myosin_N; 1.
 CC DR Pfam; PF01576; Myosin_tail; 1.
 CC DR PRINTS; PR00193; MYOSINHEAVY.
 CC DR Prodom; PD000355; myosin_head; 1.
 CC DR SMART; SM00015; IQ; 1.
 CC DR SMART; SM00242; MSC; 1.
 CC DR PROSITE; PS50096; IQ; 1.
 CC DR Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
 CC Calmodulin-binding; ATP-binding; Methylation; Multigene family;
 CC Proto-oncogene; Chromosomal translocation.
 CC KW DOMAIN 1 785 MYOSIN HEAD-LIKE.
 CC FT DOMAIN 1 786 815 IQ.
 CC FT DOMAIN 844 1934 COILED COIL (POTENTIAL).
 CC FT DOMAIN 1935 1972 CARBOXYL-TERMINAL.
 CC NP BIND 178 185 ATP (POTENTIAL).
 CC FT DOMAIN 661 683 ACTIN-BINDING (BY SIMILARITY).
 CC -----

FT DOMAIN 762 776 ACTIN-BINDING (BY SIMILARITY).
 FT MOD RES 129 129 METHYLATION (TRI-) (POTENTIAL).
 FT CONFLICT 887 889 ERK -> MSE (IN REF. 4).
 FT CONFLICT 1558 1558 T -> S (IN REF. 4).
 FT CONFLICT 1610 1611 KQ -> NE (IN REF. 4).
 FT CONFLICT 1786 1786 A -> S (IN REF. 5).
 FT CONFLICT 1958 1958 T -> L (IN REF. 4).
 SQ SEQUENCE 1972 AA; 227338 MW; 67665B2AEC61277 CRC64;

Query Match 1.1%; Score 7; DB 1; Length 1972;
 Best Local Similarity 100.0%; Pred. No. 3.5e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 94 DVASLSS 100
 DB 1309 DVASLSS 1315

RESULT 50
 TEGU_HSV6G STANDARD; PRT; 2077 AA.
 AC P30002;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Large tegument protein.
 OS U31.
 GN Human herpesvirus (type 6 / strain GS) (HHV6).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Betaherpesvirinae; Roseolovirue.
 OX NCBI_TaxID=10369;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91374623; PubMed=1654455;
 RA Josephs S.F., Ablashi D.V., Salahuddin S.Z., Jagodzinski L.L.,
 RA Wong-Staal F., Gallo R.C.;
 RT "Identification of the human herpesvirus 6 glycoprotein H and
 RT putative large tegument protein genes";
 RL J. VIROL. 65:5597-5604 (1991).
 CC -1- FUNCTION: Tegument protein.
 CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
 CC BHV-1 24, EBV BPLF1, HVS-1 64, VZV 22, AND HCMV UL48.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; S57540; AAB19786.1; -;
 CC DR PIR; H40511; W2B24.
 CC DR InterPro; IPR006928; Herpes teg_N.
 CC DR Pfam; PF04843; Herpes teg_N; 1.
 CC SQ SEQUENCE 2077 AA; 239909 MW; 28E297FDC2FAD2AE CRC64;

Query Match 1.1%; Score 7; DB 1; Length 2077;
 Best Local Similarity 100.0%; Pred. No. 3.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 601 OKRSDRA 607
 DB 691 OKRSDRA 697

Search completed: June 21, 2004, 13:27:27
 Job time : 21 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 13:23:28 ; Search time 45 Seconds

(without alignments)
4662.658 Million cell updates/sec

Title: US-10-029-345A-109

Perfect score: 665

Sequence: 1 MAHEMIGTQIVTERLVALLE.....LGVKVSQSSPSGSMETIEVS 665

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 172362

Minimum DB seq length: 473

Maximum DB seq length: 200000000

Post-processing: Listing first 75 summaries

Database :

1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*
15: SP eukaryote:*
16: SP bacteriophage:*
17: SP archaea:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	472	71.0	665	4 Q96N49	Q96N49 homo sapien
2	371	55.8	616	4 Q8N5T1	Q8N5T1 homo sapien
3	304	45.7	662	4 Q8IVT8	Q8IVT8 homo sapien
4	67	10.1	660	11 Q920R2	Q920R2 mus musculu
5	64	9.6	677	11 Q99MG6	Q99MG6 mus musculu
6	48	7.2	622	11 Q99MG5	Q99MG5 mus musculu
7	20	3.0	625	4 Q86S88	Q86S88 homo sapien
8	20	3.0	665	11 Q7TS29	Q7TS29 mus musculu
9	13	2.0	476	5 Q9VHV8	Q9VHV8 mus musculu
10	13	2.0	476	5 Q46122	Q46122 drosophila
11	9	1.4	606	5 Q8ST18	Q8ST18 caenorhabdi
12	9	1.4	657	5 Q8ST19	Q8ST19 caenorhabdi
13	8	1.2	544	16 Q826P4	Q826P4 streptomyce
14	8	1.2	559	16 Q8ABA4	Q8ABA4 bacteroides
15	8	1.2	694	16 Q8ZK16	Q8ZK16 salmonella
16	8	1.2	707	17 Q8U1R9	Q8U1R9 pyrococcus

17	8	1.2	727	16 Q8YP24	Q8YP24 anabaena sp
18	8	1.2	858	11 Q8K058	Q8K058 mus musculu
19	8	1.2	892	11 Q80T06	Q80T06 mus musculu
20	8	1.2	1001	11 Q8BMG7	Q8BMG7 mus musculu
21	8	1.2	1042	16 Q820R6	Q820R6 nitrosomona
22	8	1.2	1065	5 Q815U3	Q815U3 plasmodium
23	8	1.2	1194	10 Q9FKF1	Q9FKF1 arabidopsis
24	8	1.2	1308	5 Q01924	Q01924 caenorhabdi
25	8	1.2	1885	17 Q8TOD1	Q8TOD1 methanobact
26	7	1.1	475	16 Q51322	Q51322 borrelia bu
27	7	1.1	478	5 Q9W131	Q9W131 drosophila
28	7	1.1	479	2 Q9S3Z3	Q9S3Z3 vibrio fisc
29	7	1.1	479	10 Q9LZG8	Q9LZG8 arabidopsis
30	7	1.1	480	2 Q56817	Q56817 photorhabdu
31	7	1.1	480	5 Q8SY11	Q8SY11 drosophila
32	7	1.1	483	4 Q9HEA7	Q9HEA7 homo sapien
33	7	1.1	487	10 Q9SXY1	Q9SXY1 arabidopsis
34	7	1.1	488	16 Q8ZJT2	Q8ZJT2 yersinia pe
35	7	1.1	490	16 Q81MG2	Q81MG2 bacillus an
36	7	1.1	492	2 Q9KXD8	Q9KXD8 escherichia
37	7	1.1	494	17 Q8ZYH4	Q8ZYH4 pyrobaculum
38	7	1.1	495	5 Q9W0U3	Q9W0U3 drosophila
39	7	1.1	499	5 Q76107	Q76107 culex quin
40	7	1.1	502	5 Q9N3R4	Q9N3R4 caenorhabdi
41	7	1.1	505	12 Q806Y8	Q806Y8 cucumbe
42	7	1.1	508	16 Q7UH12	Q7UH12 rhodospir
43	7	1.1	509	2 Q50188	Q50188 mycobacteri
44	7	1.1	509	16 Q9CDE5	Q9CDE5 mycobacteri
45	7	1.1	511	16 Q8YGD2	Q8YGD2 bruceella me
46	7	1.1	511	16 Q8G1J0	Q8G1J0 bruceella me
47	7	1.1	515	2 P77815	P77815 nocardioide
48	7	1.1	516	16 Q82NP2	Q82NP2 streptomyce
49	7	1.1	519	16 Q93N44	Q93N44 coxiella bu
50	7	1.1	523	4 Q8N3D9	Q8N3D9 homo sapien
51	7	1.1	523	11 Q9CU46	Q9CU46 mus musculu
52	7	1.1	526	10 Q86WS1	Q86WS1 homo sapien
53	7	1.1	526	10 Q9ZRH9	Q9ZRH9 oryza sativ
54	7	1.1	542	12 Q39479	Q39479 avian pneum
55	7	1.1	542	12 Q39478	Q39478 avian pneum
56	7	1.1	545	16 Q8A6S1	Q8A6S1 bacteroides
57	7	1.1	546	17 Q9Y9L7	Q9Y9L7 aeropyrum p
58	7	1.1	547	5 Q816X1	Q816X1 ancylostoma
59	7	1.1	548	16 Q9TWD7	Q9TWD7 neisseria m
60	7	1.1	551	16 Q8E4X9	Q8E4X9 streptococ
61	7	1.1	551	16 Q80ZB9	Q80ZB9 streptococ
62	7	1.1	555	2 Q8KX42	Q8KX42 streptococ
63	7	1.1	555	16 Q88SR7	Q88SR7 lactobacill
64	7	1.1	564	16 Q9X144	Q9X144 thermocoga
65	7	1.1	574	11 Q8R1D2	Q8R1D2 mus musculu
66	7	1.1	576	10 Q9FLZ4	Q9FLZ4 arabidopsis
67	7	1.1	576	10 Q9SBZ2	Q9SBZ2 arabidopsis
68	7	1.1	578	16 Q9BDK0	Q9BDK0 rhizobium 1
69	7	1.1	580	16 Q8ZDF0	Q8ZDF0 streptomyce
70	7	1.1	596	4 Q9UPJ6	Q9UPJ6 homo sapien
71	7	1.1	601	16 Q86691	Q86691 streptomyce
72	7	1.1	605	6 Q7XS90	Q7XS90 sus scrofa
73	7	1.1	606	11 Q9ES58	Q9ES58 rattus norv
74	7	1.1	609	5 Q81JC3	Q81JC3 plasmodium
75	7	1.1			

ALIGNMENTS

RESULT 1
Q96N49 PRELIMINARY; PRT; 665 AA.
ID Q96N49, 01-DEC-2001 (TREMblrel. 19, Created)
AC Q96N49, 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE Hypothetical protein FLJ11411.
OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;

RP SEQUENCE FROM N.A.
RA Tashiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
RA Fukuzumi Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ichii S., Yamamoto J., Isono Y.,
RA Kawai-Hiro Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A., Sugiyama A.,
RA Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuno Y., Nagai K.,
RA Iisaga T.;

RT "NEDO human cDNA sequencing project."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AK055973; BAB71060.1; -.
DR HSSP; Q16828; IMKP.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0017017; P:MAP kinase phosphatase activity; IEA.
DR GO; GO:0006470; P:protein amino acid dephosphorylation; IEA.
DR InterPro; IPR000340; DS phosphatase.
DR InterPro; IPR001763; MAPK_phosph.
DR InterPro; IPR000387; TYR_phosphatase.
DR Pfam; PF00782; DSC; 1.
DR PRINTS; PR01764; MAPKPHPTASE.
DR SMART; SM00195; DSC; 1.
DR SMART; SM00450; RHOD; 1.
DR PROSITE; PS50206; RHODANSE_3; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS50054; TYR_PHOSPHATASE_DUAL; 1.
KW Hypothetical protein; Hydrolase.
SQ SEQUENCE 665 AA; 73058 MW; 1EA60DF08460DF CRC64;

Query Match 71.0%; Score 472; DB 4; Length 665;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 TCGKPDPIPESHFLRPVNDSECEKILPWLKSVDTIEKAKSNGCVLHCLAGISRSAT 253
DB 194 TCGKPDPIPESHFLRPVNDSECEKILPWLKSVDTIEKAKSNGCVLHCLAGISRSAT 253
QY 254 IAIAYIMKRMDSLDEAYRFVKEKRTISPNFNLGQLDYEEKIKNOTGASGPKSKL 313
DB 254 IAIAYIMKRMDSLDEAYRFVKEKRTISPNFNLGQLDYEEKIKNOTGASGPKSKL 313
QY 314 LHLKENEVPVAVSEGGQKSETPLSPCADSATSEAAQORPVHPASVPSVPSVPSLLED 373
DB 314 LHLKENEVPVAVSEGGQKSETPLSPCADSATSEAAQORPVHPASVPSVPSVPSLLED 373
QY 374 SPLVQALSGHLASADRLDSNKLKRSFLDIKSVYSASMAASLHGSSSEDALEYKPS 433
DB 374 SPLVQALSGHLASADRLDSNKLKRSFLDIKSVYSASMAASLHGSSSEDALEYKPS 433
QY 434 TTLDTGNKLCQFSVQELSEQTPETSPDKEASIPKKLOTARPSDSQSKLHSVRTSSG 493
DB 434 TTLDTGNKLCQFSVQELSEQTPETSPDKEASIPKKLOTARPSDSQSKLHSVRTSSG 493
QY 494 TAOBSLILSPHRSVGSVEDNYHTSFLGLSTSOQHLTKSAGLGLKGMHSDIILAPOTSTPSL 553
DB 494 TAOBSLILSPHRSVGSVEDNYHTSFLGLSTSOQHLTKSAGLGLKGMHSDIILAPOTSTPSL 553
QY 554 TSSWYFATSSHFSASAIYGGASVAYSCSOLPTCGDQVYVRRQKPSDRADSRSSW 613
DB 554 TSSWYFATSSHFSASAIYGGASVAYSCSOLPTCGDQVYVRRQKPSDRADSRSSW 613
QY 614 HESSEPFKQFKRRSCOMEFGESIMSENREBELGKVGSGSSFSGSMETIEVS 665
DB 614 HESSEPFKQFKRRSCOMEFGESIMSENREBELGKVGSGSSFSGSMETIEVS 665

RESULT 2
ID Q8N5T1 PRELIMINARY; PRT; 616 AA.
AC Q8N5T1;
DT 01-OCT-2002 (TREMblrel. 22, Created)
DT 01-OCT-2002 (TREMblrel. 22, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

RP SEQUENCE FROM N.A.
RC TISSUE=Brain, and Lung;
RA Strausberg R.;
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC031643; AAH31643.1; -.
DR GO; GO:0017017; P:MAP kinase phosphatase activity; IEA.
DR GO; GO:0006470; P:protein amino acid dephosphorylation; IEA.
DR InterPro; IPR000340; DS phosphatase.
DR InterPro; IPR008343; MAPK_phosph.
DR InterPro; IPR000387; TYR_phosphatase.
DR Pfam; PF00782; DSC; 1.
DR PRINTS; PR01764; MAPKPHPTASE.
DR SMART; SM00195; DSC; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS50054; TYR_PHOSPHATASE_DUAL; 1.
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 616 AA; 67636 MW; 2CB0B14482F2AD72 CRC64;

Query Match 55.8%; Score 371; DB 4; Length 616;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 471; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 194 TCGKPDPIPESHFLRPVNDSECEKILPWLKSVDTIEKAKSNGCVLHCLAGISRSAT 253
DB 145 TCGKPDPIPESHFLRPVNDSECEKILPWLKSVDTIEKAKSNGCVLHCLAGISRSAT 204
QY 254 IAIAYIMKRMDSLDEAYRFVKEKRTISPNFNLGQLDYEEKIKNOTGASGPKSKL 313
DB 205 IAIAYIMKRMDSLDEAYRFVKEKRTISPNFNLGQLDYEEKIKNOTGASGPKSKL 264
QY 314 LHLKENEVPVAVSEGGQKSETPLSPCADSATSEAAQORPVHPASVPSVPSVPSLLED 373
DB 265 LHLKENEVPVAVSEGGQKSETPLSPCADSATSEAAQORPVHPASVPSVPSVPSLLED 324
QY 374 SPLVQALSGHLASADRLDSNKLKRSFLDIKSVYSASMAASLHGSSSEDALEYKPS 433
DB 325 SPLVQALSGHLASADRLDSNKLKRSFLDIKSVYSASMAASLHGSSSEDALEYKPS 384
QY 434 TTLDTGNKLCQFSVQELSEQTPETSPDKEASIPKKLOTARPSDSQSKLHSVRTSSG 493
DB 385 TTLDTGNKLCQFSVQELSEQTPETSPDKEASIPKKLOTARPSDSQSKLHSVRTSSG 444
QY 494 TAOBSLILSPHRSVGSVEDNYHTSFLGLSTSOQHLTKSAGLGLKGMHSDIILAPOTSTPSL 553
DB 445 TAOBSLILSPHRSVGSVEDNYHTSFLGLSTSOQHLTKSAGLGLKGMHSDIILAPOTSTPSL 504
QY 554 TSSWYFATSSHFSASAIYGGASVAYSCSOLPTCGDQVYVRRQKPSDRADSRSSW 613
DB 505 TSSWYFATSSHFSASAIYGGASVAYSCSOLPTCGDQVYVRRQKPSDRADSRSSW 564
QY 614 HESSEPFKQFKRRSCOMEFGESIMSENREBELGKVGSGSSFSGSMETIEVS 665
DB 565 HESSEPFKQFKRRSCOMEFGESIMSENREBELGKVGSGSSFSGSMETIEVS 616

RESULT 3
ID Q81VT8 PRELIMINARY; PRT; 662 AA.
ID Q81VT8

AC Q81V78; (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Similar to dual specificity phosphatase 16.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 RX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Dundenum;
 RA Strausberg R.;
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC042101; AAH42101.1; -
 DR GO; GO:0017017; F:MAP kinase phosphatase activity; IEA.
 DR GO; GO:0006470; P:Protein amino acid dephosphorylation; IEA.
 DR InterPro; IPR000340; DS phosphatase.
 DR InterPro; IPR008343; MAPK phosph.
 DR InterPro; IPR001763; Rhodanese-like.
 DR InterPro; IPR000387; TYR_phosphatase.
 DR Pfam; PF00782; DSpC; 1.
 DR Pfam; PF00581; Rhodanese; 1.
 DR PRINTS; PR01764; MAPKDPHTASE.
 DR SMART; SM00195; DSpC; 1.
 DR SMART; SM00450; RHOD; 1.
 DR PROSITE; PSS0206; RHODANSE_3; 1.
 DR PROSITE; PSS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE; PSS0056; TYR_PHOSPHATASE_2; 1.
 DR PROSITE; PSS0054; TYR_PHOSPHATASE_DUAL; 1.
 DR PROSITE; PSS0054; TYR_PHOSPHATASE_DUAL; 1.
 SQ SEQUENCE 662 AA; 72818 MW; 350534EF0652B98F CRC64;

Query Match 45.7%; Score 304; DB 4; Length 662;
 Best Local Similarity 100.0%; Pred. No. 3e-313;
 Matches 304; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 362 SVSPVPSLLEDSPLVQALSGLHLSADRLSDSNLKRSPFLDIKSVSYASMAASLHGFS 421
 DB 359 SVSPVPSLLEDSPLVQALSGLHLSADRLSDSNLKRSPFLDIKSVSYASMAASLHGFS 418
 QY 422 SSEDALERYKPSSTLLDGTNKLCOFSPVQELSEQTPETSPDKEASIPKLTQTARPSDSQ 481
 DB 419 SSEDALERYKPSSTLLDGTNKLCOFSPVQELSEQTPETSPDKEASIPKLTQTARPSDSQ 478
 QY 482 KRLHSVRTSSSGTGAQRSLSPHRSQSVENDYHTSPFGISTGQHLTKSAGLGLKQMS 541
 DB 479 KRLHSVRTSSSGTGAQRSLSPHRSQSVENDYHTSPFGISTGQHLTKSAGLGLKQMS 538
 QY 542 DILAPOTSTSLTSSWFATSSSHFYASAIYGSASYSAYSCSOLPTGDOYVSVRRQ 601
 DB 539 DILAPOTSTSLTSSWFATSSSHFYASAIYGSASYSAYSCSOLPTGDOYVSVRRQ 598
 QY 602 KPDRADSRSSWTEESPFKQFKRSCQMEFGESINSRRELGLGVSGSSFFSGSMET 661
 DB 599 KPDRADSRSSWTEESPFKQFKRSCQMEFGESINSRRELGLGVSGSSFFSGSMET 658
 QY 662 IEVS 665
 DB 659 IEVS 662

RESULT 4
 Q920R2 PRELIMINARY; PRT; 660 AA.
 ID Q920R2
 AC Q920R2; (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE MAP kinase phosphatase-7.
 GN DUSP16 OR 3830417M17RIK OR MKP-7.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Cerebellum;
 RX MEDLINE=21486429; PubMed=11489891;
 RA Masuda K., Shima H., Watanabe M., Kikuchi K.;
 RT "MKP-7, a Novel Mitogen-activated Protein Kinase Phosphatase,
 Functions as a Shuttle Protein";
 RL J. Biol. Chem. 276:39002-39011(2001).
 DR EMBL; AB052157; BAB47240.1; -
 DR HSSP; Q16828; IMKP.
 DR MGD; MG1:1917936; Dusp16.
 DR GO; GO:0005737; C:cytoplasm; IDA.
 DR GO; GO:0005634; C:nucleus; IDA.
 DR GO; GO:0005515; F:protein binding; IPI.
 DR GO; GO:0001088; P:inactivation of MAPK; IDA.
 DR InterPro; IPR000340; DS phosphatase.
 DR InterPro; IPR008343; MAPK phosph.
 DR InterPro; IPR001763; Rhodanese-like.
 DR InterPro; IPR000387; TYR_phosphatase.
 DR Pfam; PF00782; DSpC; 1.
 DR Pfam; PF00581; Rhodanese; 1.
 DR PRINTS; PR01764; MAPKDPHTASE.
 DR SMART; SM00195; DSpC; 1.
 DR SMART; SM00450; RHOD; 1.
 DR PROSITE; PSS0206; RHODANSE_3; 1.
 DR PROSITE; PSS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE; PSS0056; TYR_PHOSPHATASE_2; 1.
 DR PROSITE; PSS0054; TYR_PHOSPHATASE_DUAL; 1.
 DR PROSITE; PSS0054; TYR_PHOSPHATASE_DUAL; 1.
 KM Hydrolyase; Kinase.
 SQ SEQUENCE 660 AA; 72695 MW; DB609FCDAD4AA309 CRC64;

Query Match 10.1%; Score 67; DB 11; Length 660;
 Best Local Similarity 100.0%; Pred. No. 1.5e-61;
 Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 LVALLSGREKVLILSRPVEYNTSHILEAININSGKMKRLQDQKVLITELIHSK 74
 DB 15 LVALLSGREKVLILSRPVEYNTSHILEAININSGKMKRLQDQKVLITELIHSK 74
 QY 75 HKVDIDC 81
 DB 75 HKVDIDC 81

RESULT 5
 Q99MG6 PRELIMINARY; PRT; 677 AA.
 ID Q99MG6
 AC Q99MG6; (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
 DE Map kinase phosphatase-M A1 isoform.
 GN DUSP16 OR 3830417M17RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/c;
 RA Matsuguchi T., Musikacharoen T., Johnson T.R., Kraft A.S.,
 RA Yoshikai Y.;
 RT "A Novel MKP Is an Important Negative Regulator of LPS-mediated JNK
 Activation in Macrophages";
 RL Mol. Cell. Biol. 20:6999-7009(2001).
 DR EMBL; AF345951; AAK35052.1; -
 DR HSSP; Q16828; IMKP.
 DR MGD; MG1:1917936; Dusp16.
 DR GO; GO:0005737; C:cytoplasm; IDA.
 DR GO; GO:0005634; C:nucleus; IDA.
 DR GO; GO:0005515; F:protein binding; IPI.
 DR GO; GO:0001088; P:inactivation of MAPK; IDA.

DR InterPro; IPR000340; DS phosphatase.
 DR InterPro; IPR008343; MAPK phosph.
 DR InterPro; IPR001763; Rhodanese-like.
 DR InterPro; IPR000387; Tyr_phosphatase.
 DR Pfam; PF00782; DSpC; 1.
 DR Pfam; PF00581; Rhodanese; 1.
 DR PRINTS; PRO1764; MAPKPHPTASE.
 DR SMART; SM00195; DSpC; 1.
 DR SMART; SM00450; RHOD; 1.
 DR PROSITE; PS50206; RHODANESE_3; 1.
 DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
 DR PROSITE; PS50054; TYR_PHOSPHATASE_DUAL; 1.
 KW Hydrolyase; Kinase.
 SQ SEQUENCE 677 AA; 74550 MW; 8B6DSB7096CCE2FC CRC64;

Query Match
 Best Local Similarity 100.0%; Pred. No. 2,3e-58;
 Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 SFNSVHLIAGGFAEFSCFPGLCGKSTIVPTICISOPCLPVANIGPRLIPNLVLCQORD 173
 |||||
 DB 114 SFNSVHLIAGGFAEFSCFPGLCGKSTIVPTICISOPCLPVANIGPRLIPNLVLCQORD 173

QY 174 VLNK 177
 |||||
 DB 174 VLNK 177

RESULT 6
 ID Q99MG5 PRELIMINARY; PRT; 622 AA.
 AC Q99MG5;
 DT 01-JUN-2001 (TrEMBLrel. 1.7, Created)
 DT 01-JUN-2001 (TrEMBLrel. 1.7, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Map kinase phosphatase-M A2 isoform.
 GN DUSP16 OR 3830417ML7RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCB1_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/C;
 RA Matsuguchi T., Musikachareon T., Johnson T.R., Kraft A.S.,
 RA Yoshikai Y.;
 RT "A Novel MKR is an Important Negative Regulator of LPS-mediated JNK
 RT Activation in Macrophages.";
 RL Mol. Cell. Biol. 20:6999-7009(2001).
 DR EMBL; AF345952; AAK35053.1; -.
 DR HSSP; Q16828; IMKP.
 DR MGD; MGI:1917936; Dusp16.
 DR GO; GO:0005737; Cytoplasm; IDA.
 DR GO; GO:0005634; Cytosol; IDA.
 DR GO; GO:0005515; Protein binding; IPI.
 DR GO; GO:000188; P-inactivation of MAPK; IDA.
 DR InterPro; IPR000340; DS phosphatase.
 DR InterPro; IPR008343; MAPK_phosph.
 DR InterPro; IPR001763; Rhodanese-like.
 DR InterPro; IPR000387; Tyr_phosphatase.
 DR Pfam; PF00782; DSpC; 1.
 DR Pfam; PF00581; Rhodanese; 1.
 DR PRINTS; PRO1764; MAPKPHPTASE.
 DR SMART; SM00195; DSpC; 1.
 DR SMART; SM00450; RHOD; 1.
 DR PROSITE; PS50206; RHODANESE_3; 1.
 DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
 DR PROSITE; PS50054; TYR_PHOSPHATASE_DUAL; 1.
 KW Hydrolyase; Kinase.
 SQ SEQUENCE 622 AA; 68672 MW; 6CCCAAE6B9099B98 CRC64;

Query Match
 Best Local Similarity 7.2%; Score 48; DB 11; Length 622;
 Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 TCPKDFIPESHFLRPVNDSCFKILPWLKSYDVFIEKAKASNCVL 241
 |||||
 DB 139 TCPKDFIPESHFLRPVNDSCFKILPWLKSYDVFIEKAKASNCVL 186

RESULT 7
 ID Q86S88 PRELIMINARY; PRT; 625 AA.
 AC Q86S88;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-OCT-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Similar to dual specificity phosphatase 8.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 NC NCB1_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain, and Astrocytoma;
 RA Strausberg R.;
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC045110; AAK45110.1; -.
 DR GO; GO:0017017; F:MAP kinase phosphatase activity; IEA.
 DR GO; GO:0006470; P:protein amino acid dephosphorylation; IEA.
 DR InterPro; IPR000340; DS phosphatase.
 DR InterPro; IPR008343; MAPK_phosph.
 DR InterPro; IPR002965; P-rich_extensn.
 DR InterPro; IPR001763; Rhodanese-like.
 DR InterPro; IPR000387; Tyr_phosphatase.
 DR Pfam; PF00782; DSpC; 1.
 DR PRINTS; PRO1217; PRICEXTENSN.
 DR SMART; SM00195; DSpC; 1.
 DR SMART; SM00450; RHOD; 1.
 DR PROSITE; PS50206; RHODANESE_3; 1.
 DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
 DR PROSITE; PS50054; TYR_PHOSPHATASE_DUAL; 1.
 SQ SEQUENCE 625 AA; 65826 MW; C7C808407B724FFC CRC64;

Query Match
 Best Local Similarity 3.0%; Score 20; DB 4; Length 625;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 242 VHCLAGISRSATIAIAYIMK 261
 |||||
 DB 244 VHCLAGISRSATIAIAYIMK 263

RESULT 8
 ID Q7TS29 PRELIMINARY; PRT; 665 AA.
 AC Q7TS29;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Dusp8 protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCB1_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Brain;
 RX MEDLINE=22386257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,
 RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schaefer T.E.,
 RA Brownstein M.J., Udin T.B., Toehiyki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
 RA Bosak S.A., McKean P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fehy J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bonfield G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
 RA Krzyzanski M.I., Skalka U., Smalls D.E., Schnerch A., Schein J.E.,
 RA Jones S.J., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length human
 RT and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 [12]
 RP SEQUENCE FROM N.A.
 RA STRAIN=C57BL/6; TISSUE=Brain;
 RC Strausberg R.;
 RL Submitted (May-2003) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC052705; AAH52705.1; -
 SQ SEQUENCE 665 AA; 65021 MW; 9166E36A835249F CRC64;

Query Match 3.0%; Score 20; DB 11; Length 665;
 Best Local Similarity 100.0%; Pred. No. 1,2e-11;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 242 VHCLAGSRATTAIAYIMK 261
 |||||
 Db 244 VHCLAGSRATTAIAYIMK 263

RESULT 9
 Q9YVH8 PRELIMINARY; PRT; 476 AA.
 AC Q9YVH8;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE PUC protein (SD08157P).
 DE PUC OR CG7850.
 OS Drosophila melanogaster (Fruit fly)
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NX NCBI_TaxID=7227;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkely;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananidis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wotman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Peiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Berno P.V., Berman B.P., Bhandari D., Boltekar S.,
 RA Borkova D., Botchan M.R., Bouck J., Brockstein P., Brotter P.,
 RA Burks K.C., Busam D.A., Butler H., Cadieu E., Cantor A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Houten J.D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalish F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Matrei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Mikhlov G., Milshina N.V., Moberly C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Murry D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spralling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Waasman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 [12]
 RP SEQUENCE FROM N.A.
 RA Stapleton M., Brockstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise B., George R.,
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
 RA Nuno J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AE003677; AAF54191.1; -
 DR EMBL: AY061616; AAL29164.1; -
 DR HSSP: Q16828; IMKP.
 DR FlyBase: FBgn004210; puc.
 DR GO: GO:0006579; P:JUN kinase phosphatase activity; IMP.
 DR GO: GO:0017017; P:MAP kinase phosphatase activity; NAS.
 DR GO: GO:0030036; P:actin cytoskeleton organization and biogenesis; IMP.
 DR GO: GO:0046843; P:dorsal appendage formation; IMP.
 DR GO: GO:0008544; P:epidermal differentiation; IMP.
 DR GO: GO:0007254; P:JNK cascade; IMP.
 DR GO: GO:0046844; P:mitogen-activated protein kinase cascade; NAS.
 DR GO: GO:0046329; P:negative regulation of JNK cascade; NAS.
 DR GO: GO:0016318; P:prometastatic rotation; NAS.
 DR GO: GO:0006470; P:protein amino acid dephosphorylation; IDA.
 DR GO: GO:0007396; P:protein amino acid dephosphorylation; NAS.
 DR GO: GO:0042060; P:protein dephosphorylation; NAS.
 DR InterPro: IPR000340; DS phosphatase.
 DR InterPro: IPR000387; Tyr_phosphatase.
 DR Pfam: PF00782; DSPC; 1.
 DR SMART: SM00193; DSPC; 1.
 DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE: PS50056; TYR_PHOSPHATASE_2; 1.
 DR PROSITE: PS50054; TYR_PHOSPHATASE_DUAL; 1.
 KW Hydrolyase.
 SQ SEQUENCE 476 AA; 51270 MW; 8E5DC281349F0F4F CRC64;

Query Match 2.0%; Score 13; DB 5; Length 476;
 Best Local Similarity 100.0%; Pred. No. 0.00024;
 Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 246 AGISRSATTAIAY 258
 |||||
 Db 218 AGISRSATTAIAY 230

RESULT 10
 O46122 PRELIMINARY; PRT; 476 AA.
 ID O46122;
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE PUCKERED protein.
 DE PUC OR PUCKERED OR CG7850.
 OS Drosophila melanogaster (Fruit fly)
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.

DR PROSITE; PSS0056; TYR_PHOSPHATASE_2; 1.
 DR PROSITE; PSS0054; TYR_PHOSPHATASE_DUAL; 1.
 KM Hypothetical protein; Hydrolase.
 SQ SEQUENCE 657 AA; 71002 MW; A0D953DB6326B43 CRC64;

Query Match 1.4%; Score 9; DB 5; Length 657;
 Best Local Similarity 100.0%; Pred. No. 5.7;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 243 HCLAGISRS 251
 |||||
 DB 261 HCLAGISRS 269

RESULT 13

ID Q82GP4 PRELIMINARY; PRT; 544 AA.
 AC Q82GP4;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Putative integral membrane transport protein.
 GN SAV3853.
 OS Streptomyces avermectilis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_Taxid=33903;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=21477403; PubMed=11572948;
 RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
 RA Shiiose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
 RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
 RT "Genome sequence of an industrial microorganism Streptomyces
 RT avermectilis: deducing the ability of producing secondary
 RT metabolites.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).

RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
 RX MEDLINE=22608306; PubMed=12692562;
 RA Ikeda H., Ishikawa J., Hanamoto A., Shiiose M., Kikuchi H., Shiba T.,
 RA Sakaki Y., Hattori M., Omura S.;
 RT "Complete genome sequence and comparative analysis of the industrial
 RT microorganism Streptomyces avermectilis.";
 RL Nat. Biotechnol. 21:526-531(2003).
 DR EMBL; AP005036; BAC71565.1; -
 KM Complete proteome.
 SQ SEQUENCE 544 AA; 56291 MW; FC13A5045264A9DA CRC64;

Query Match 1.2%; Score 8; DB 16; Length 544;
 Best Local Similarity 100.0%; Pred. No. 56;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 548 TSPSRLTS 555
 |||||
 DB 13 TSPSRLTS 20

RESULT 14

ID Q8AAB4 PRELIMINARY; PRT; 559 AA.
 AC Q8AAB4;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Asparagine synthetase B.
 GN B70551.
 OS Bacteroides thetaiotaomicron.
 OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;
 OC Bacteroidaceae; Bacteroides.
 OX NCBI_Taxid=818;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=VPI-5482 / ATCC 29148;
 RX MEDLINE=22550858; PubMed=12663928;
 RA Xu J., Bjurell M.K., Himrod J., Deng S., Carmichael L.K.,
 RA Chiang H.C., Hooper L.V., Gordon J.I.;
 RT "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis.";
 RL Science 299:2074-2076(2003).
 DR EMBL; A6016928; AAO75658.1; -
 DR GO; GO:0004066; F:asparagine synthase (glutamine-hydrolyzing) . . .; IEA.
 DR GO; GO:0006529; P:asparagine biosynthesis; IEA.
 DR GO; GO:0008152; P:metabolism; IEA.
 DR InterPro; IPR001962; Asn_synthase.
 DR InterPro; IPR006426; Asn_synth_AEB.
 DR Pfam; PF00733; Asn_synthase_1.
 DR Pfam; PF00310; GATase_2; 1.
 DR TIGRPFAM; TIGR01536; asn_synth_AEB; 1.
 KM Complete proteome.
 SQ SEQUENCE 559 AA; 63019 MW; 292F181736AD4B05 CRC64;

Query Match 1.2%; Score 8; DB 16; Length 559;
 Best Local Similarity 100.0%; Pred. No. 57;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 359 SVPSVPSV 366
 |||||
 DB 514 SVPSVPSV 521

RESULT 15

ID Q8ZK16 PRELIMINARY; PRT; 694 AA.
 AC Q8ZK16;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Putative ATP-dependent Lon protease.
 GN STM4491.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 OX NCBI_Taxid=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=L72 / SGGC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du P., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewel N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 RT LT2.";
 RL Nature 413:852-856(2001).

DR EMBL; A6008811; AA123309.1; -
 DR GO; GO:0008233; F:peptidase activity; IEA.
 KM Protease; Hypothetical protein; Complete proteome.
 SQ SEQUENCE 694 AA; 77463 MW; 53BE7B95B5F8B878 CRC64;

Query Match 1.2%; Score 8; DB 16; Length 694;
 Best Local Similarity 100.0%; Pred. No. 70;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 322 PVPVASEG 329
 |||||

DB 8 PVPVASEG 15

RESULT 16

ID Q8UR19 PRELIMINARY; PRT; 707 AA.
 AC Q8UR19;

DT 01-JUN-2002 (TReMBLrel. 21, Created)
 DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Hypothetical protein P1136.
 GN P1136.
 OS *Pyrococcus furiosus*.
 OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
 OC Pyrococcus.
 OX NCBI_TaxID=2261;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Vcl / DSM 3638 / ATCC 43587 / JCM 8422;
 RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
 RT "The complete sequence of the *Pyrococcus furiosus* genome."
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF010223; AAL81260.1; -.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 707 AA; 80090 MW; 012ACBC9C987A6D7 CRC64;

Query Match 1.2%; Score 8; DB 17; Length 707;
 Best Local Similarity 100.0%; Pred. No. 71;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 SGTETKVL 28
 |||||
 DB 53 SGTETKVL 60

RESULT 17

ID 08YP24 PRELIMINARY; PRT; 727 AA.
 AC 08YP24;
 DT 01-MAR-2002 (TReMBLrel. 20, Created)
 DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
 DE Hypothetical protein A114378.
 GN A114378.
 OS *Anabaena* sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 OX NCBI_TaxID=103690;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=11595285; PubMed=11759840;
 RA Kaneo T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
 RA Watanabe A., Iriyuchi M., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
 RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
 RA Yasuda M., Tabata S.;
 RT "Complete genomic sequence of the filamentous nitrogen-fixing
 cyanobacterium *Anabaena* sp. strain PCC 7120."
 RL DNA Res. 8:205-213 (2001).
 DR EMBL; AP003596; BAB76077.1; -.
 DR PIR; AB2353; AB2353.
 DR InterPro; IPR008941; TPR-like.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 727 AA; 81109 MW; 78E7D9B9FA603183 CRC64;

Query Match 1.2%; Score 8; DB 16; Length 727;
 Best Local Similarity 100.0%; Pred. No. 72;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 359 SVPSVPSV 366
 |||||
 DB 116 SVPSVPSV 123

RESULT 18

ID 08K058 PRELIMINARY; PRT; 858 AA.
 AC 08K058;
 DT 01-OCT-2002 (TReMBLrel. 22, Created)
 DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
 DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)

DE Hypothetical protein (Fragment).

OS *Mus musculus* (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Retina;
 RA Strausberg R.;
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC034090; AAH34090.1; -.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR InterPro; IPR001478; PDZ.
 DR Pfam; PF00595; PDZ; 1.
 DR PROSITE; PS50106; PDZ; 1.
 KW Hypothetical protein.
 FT NON TER 1
 SQ SEQUENCE 858 AA; 92849 MW; AAE908E3FA574ED CRC64;

Query Match 1.2%; Score 8; DB 11; Length 858;
 Best Local Similarity 100.0%; Pred. No. 84;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 544 LAPQSTP 551
 |||||
 DB 670 LAPQSTP 677

RESULT 19

ID 080T06 PRELIMINARY; PRT; 892 AA.
 AC 080T06;
 DT 01-JUN-2003 (TReMBLrel. 24, Created)
 DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
 DE MK1A0839 protein (Fragment).
 GN MK1A0839.
 OS *Mus musculus* (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX TISSUE=Brain;
 RX MEDLINE=22579291; PubMed=12693553;
 RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
 RA Nakajima D., Nagase T., Ohara O., Koga H.;
 RT "Prediction of the coding sequences of mouse homologues of KIA gene:
 RT II. The complete nucleotide sequences of 400 mouse KIAA-homologous
 RT cDNAs identified by screening of terminal sequences of cDNA clones
 randomly sampled from size-fractionated libraries."
 RL DNA Res. 10:35-48 (2003).
 DR EMBL; AK12385; BAC65667.2; -.
 DR NON TER 1
 FT NON TER 1
 FT NON TER 892
 SQ SEQUENCE 892 AA; 99707 MW; F2204825B83130A9 CRC64;

Query Match 1.2%; Score 8; DB 11; Length 892;
 Best Local Similarity 100.0%; Pred. No. 87;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 309 SKLKLHL 316
 |||||
 DB 205 SKLKLHL 212

RESULT 20

ID 08BMG7 PRELIMINARY; PRT; 1001 AA.
 AC 08BMG7;
 DT 01-MAR-2003 (TReMBLrel. 23, Created)
 DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)

DE Similar to RAB3-GAP regulatory domain.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCBL_TaxID=10090;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Forelimb;
 RX MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium.
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 60,770 full-length cDNAs."
 RL Nature 420:563-573(2002).
 DR EMBL: AK031191; BAC27296.1; -;
 SQ SEQUENCE 1001 AA; 112208 MW; BCB82508E8201A63 CRC64;

Query Match 1.2%; Score 8; DB 11; Length 1001;
 Best Local Similarity 100.0%; Pred. No. 97;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 309 SKIKLHL 316
 |||||
 DB 271 SKIKLHL 278

RESULT 21

ID Q820R6 PRELIMINARY; PRT; 1042 AA.
 AC Q820R6;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Acriflavin resistance protein:heavy metal efflux pump CzCA.
 GN NE0345.
 OS Nitrosomonas europaea.
 OC Bacteria; Proteobacteria; Betaproteobacteria; Nitrosomonadales;
 OC Nitrosomonadaceae; Nitrosomonas.
 NC NCBL_TaxID=915;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 19718 / IFO 14298;
 RX MEDLINE=22586410; PubMed=12700255;
 RA Chain P., Lamerdin J.E., Larimer F.W., Regala W., Lao V., Land M.,
 Hauser L., Hooper A.B., Klotz M.G., Norton J., Sayavedra-Soto L.A.,
 Aciego D.M., Hommes N.G., Whitaker M.W., Arp D.J.,
 "Complete genome sequence of the ammonia-oxidizing bacterium and
 RT obligate chemolithoautotroph Nitrosomonas europaea."
 RL J. Bacteriol. 185:2759-2773(2003).
 DR EMBL: BX321857; CAD84256.1; -;
 DR GO: GO:0016021; C: integral to membrane; IEA.
 DR GO: GO:0008324; F: cation transporter activity; IEA.
 DR GO: GO:0006812; P: cation transport; IEA.
 DR InterPro: IPR001036; Acrflavin_re.
 DR InterPro: IPR004763; CzCA.
 DR Pfam: PF00873; ACR_tran; 1.
 DR PRINTS: PR00702; ACRIFLAVINRP.
 DR TIGRFAMs: TIGR00914; 2A0601; 1.
 KM Complete proteome.
 SQ SEQUENCE 1042 AA; 112982 MW; 471082BEC2F6B5E6 CRC64;

Query Match 1.2%; Score 8; DB 16; Length 1042;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 379 ALSGLHS 386
 |||||
 DB 923 ALSGLHS 930

RESULT 22
 Q815U3 PRELIMINARY; PRT; 1065 AA.
 ID Q815U3

AC Q815U3;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein.
 GN PF10530C.
 OS Plasmodium falciparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 NC NCBL_TaxID=36329;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=3D7;
 RX MEDLINE=22255705; PubMed=12368864;
 RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
 RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
 RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,
 RA Chan M.-S., Nene V., Shallow S.J., Sun B., Peterson J., Angiuoli S.,
 RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,
 RA Martin D.M.A., Fairclimb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,
 RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
 RA Venter J.C., Carnucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
 RA Fraser C.M., Barrall B.,
 RT "Genome sequence of the human malaria parasite Plasmodium
 falciparum."
 RL Nature 419:498-511(2002).
 DR EMBL: AE014845; AAN36195.1; -;
 KM Hypothetical protein.
 SQ SEQUENCE 1065 AA; 122856 MW; 0270CD53F55CF40 CRC64;

Query Match 1.2%; Score 8; DB 5; Length 1065;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 549 STPSLTSS 556
 |||||
 DB 325 STPSLTSS 332

RESULT 23

ID Q9FKF1 PRELIMINARY; PRT; 1194 AA.
 AC Q9FKF1;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE DNA-directed RNA polymerase subunit (EC 2.7.7.6) (DNA-directed RNA
 polymerase beta chain).
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC Eurosid II; Brassicales; Brassicaceae; Arabidopsia.
 NC NCBL_TaxID=3702;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=Columbia;
 RX MEDLINE=98403884; PubMed=9734815;
 RA Kocani H., Nakamura Y., Sato S., Asamizu E., Kaneko T., Miyajima N.,
 RA Tabata S.,
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. VI.
 RT Physically assigned pi and TAC clones."
 RL DNA Res. 5:203-216(1998).
 CC -1- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N DIPHOSPHATE +
 CC {RNA}(N).
 CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.
 DR EMBL: AB012240; BAB1387.1; -;
 DR GO: GO:0003677; F: DNA binding; IEA.
 DR GO: GO:0003899; F: DNA-directed RNA polymerase activity; IEA.
 DR GO: GO:0003900; F: DNA-directed RNA polymerase I activity; IEA.
 DR GO: GO:0003901; F: DNA-directed RNA polymerase II activity; IEA.
 DR GO: GO:0003902; F: DNA-directed RNA polymerase III activity; IEA.
 DR GO: GO:0016740; F: transferase activity; IEA.
 DR GO: GO:0006350; P: transcription; IEA.

```
DR InterPro; IPR007121; RNA_pol_B.
DR InterPro; IPR007644; RNA_pol_Rpb2_1.
DR InterPro; IPR007642; RNA_pol_Rpb2_2.
DR InterPro; IPR007645; RNA_pol_Rpb2_3.
DR InterPro; IPR007646; RNA_pol_Rpb2_4.
DR InterPro; IPR007647; RNA_pol_Rpb2_5.
DR InterPro; IPR007120; RNA_pol_Rpb2_6.
DR InterPro; IPR007641; RNA_pol_Rpb2_7.
DR Pfam; PF04563; RNA_pol_Rpb2_1.
DR Pfam; PF04561; RNA_pol_Rpb2_2.
DR Pfam; PF04565; RNA_pol_Rpb2_3.
DR Pfam; PF04566; RNA_pol_Rpb2_4.
DR Pfam; PF04567; RNA_pol_Rpb2_5.
DR Pfam; PF00562; RNA_pol_Rpb2_6.
DR Pfam; PF04560; RNA_pol_Rpb2_7.
DR PROSITE; PS01166; RNA_POL_BETA.1.
SQ DNA-directed RNA polymerase; Transcription; Transferase.
SEQUENCE 1194 AA; 134102 MW; CD48F69D16C50F8E CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 10; Length 1194;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 22 GTEKVLII 29
Db 203 GTEKVLII 210

RESULT 24
O01924 PRELIMINARY; PRT; 1308 AA.
ID 001924;
AC 001924;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUN-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE R155.2 protein.
GN R155.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2.
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2.
RA Geisel C., Wamsley P., Kramer J.;
RT "The sequence of C. elegans cosmid R155.";
RL submitted (May-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF003390; AAB54272.1; -.
DR PIR; T15280; T15280.
DR WormPep; R155.2; CE12864.
DR GO; GO:0004725; F:protein tyrosine phosphatase activity; IEA.
DR GO; GO:0006470; P:protein amino acid dephosphorylation; IEA.
DR InterPro; IPR000387; TYR_phosphatase.
DR InterPro; IPR000242; TYR_PP.
DR InterPro; IPR003125; MSN.
DR Pfam; PF02206; MSN.1.
DR Pfam; PF00102; Y_phosphatase; 1.
DR SMART; SM00194; PTPC; 1.
DR SMART; SM00453; MSN; 1.
DR PROSITE; PSS0056; TYR_PHOSPHATASE_2; 1.

DR PROSITE; PSS0055; TYR_PHOSPHATASE_PTP; 1.
SQ SEQUENCE 1308 AA; 146780 MW; 8FEEA85099312D CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 5; Length 1308;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 247 GISRSATI 254
Db 1165 GISRSATI 1172

RESULT 25
O08TOD1 PRELIMINARY; PRT; 1885 AA.
ID O08TOD1;
AC O08TOD1;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE TPR-domain containing protein.
DE MALC13.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Euryarchaeota orders incertae sedis;
OC Methanosarcinales; Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932238;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., MacDonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Atnoor D., Brown A.,
RA Allen N., Naylor J., Strange-Thomann N., DeBrellano K., Johnson R.,
RA Linton L., McSwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grathame D.A., Guss A.M.,
RA Hedderich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
and physiological diversity.";
RL Genome Res. 12:532-542(2002).
DR EMBL; AE010832; AAM05026.1; -.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003676; F:nucleic acid binding; IEA.
DR GO; GO:0006396; P:RNA processing; IEA.
DR InterPro; IPR001064; Crystallin.
DR InterPro; IPR003107; HAT.
DR InterPro; IPR000504; RNA_rec_mot.
DR InterPro; IPR001440; TPR.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF00515; TPR; 50.
DR SMART; SM00386; HAT; 12.
DR SMART; SM00028; TPR; 48.
DR PROSITE; PSS00225; CRYSTALLIN_BETAGAMMA.
DR PROSITE; PS00030; RRM_RNP_1; 1.
DR KW Complete proteome.
SQ SEQUENCE 1885 AA; 214339 MW; 6D15283773777AA2 CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 17; Length 1885;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 106 VLGLTKER 113
Db 1840 VLGLTKER 1847

RESULT 26
O051322 PRELIMINARY; PRT; 475 AA.
ID O051322;
AC O051322;
```

DT 01-JUN-1998 (TrEMBLrel. 06, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Fibronectin/fibronogen-binding protein, putative.
 OS BB0347.
 OS Borrelia burgdorferi (Lyme disease spirochete).
 OC Bacteria; Spirochaetes; Spirochaetales; Spirochaetaceae; Borrelia.
 NCBI_TaxId=139;
 [1]
 RP SEQUENCE FROM N.A.
 RC SRRAIN=ATCC 35210 / B31;
 RA MEDLINE=9403685;
 RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
 RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
 RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
 RA Peterson J., Karpavavage A.R., Quackenbush J., Salzberg S., Hanson M.,
 RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
 RA Uetebach T., Wathley L., McDonald L., Artiach P., Bowman C.,
 RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
 RA Smith H.O., Venter J.C.;
 RT "Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi";
 RL Nature 390:580-586(1997).
 DR EMBL; AB001141; AAC6734.1; -.
 DR PIR; B70143; B70143.
 DR TIGR; BB0347; -.
 DR InterPro; IPR006532; DUF814.
 DR InterPro; IPR008616; Fbpa.
 DR Pfam; PF05670; DUF814; 1.
 DR Pfam; PF05833; Fbpa; 1.
 KW Complete proteome.
 SQ SEQUENCE 475 AA; 55912 MW; 94E07528DB6EB4 CRC64;
 Query Match 1.1%; Score 7; DB 16; Length 475;
 Best Local Similarity 100.0%; Pred. No. 5.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 296 KKKINOT 302
 Db 469 KKKINOT 475
 RESULT 27
 Q9W131
 ID 09W131 PRELIMINARY; PRT; 478 AA.
 AC Q9W131;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE CG13594 protein.
 GN CG13594.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxId=7227;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RA MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abri J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballow R.M., Basu A., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz S., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck C.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasbo P., Lei Y., Levitky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milenina N.V., Modyarty C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Munzy D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusslein D.R., Pacle J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Sanders R.D.C., Scheller F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Weissman D.A., Weinstein G.M., Weisenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195 (2000).
 [2]
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorese V., Doup L.E., Doyle C., Drenek D., Fartan D.,
 RA Ferreira S., Friese E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacle J., Paragay V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheller F.,
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of Drosophila melanogaster genome";
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 [3]
 RP SEQUENCE FROM N.A.
 RA Mista S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Friese E., de Grey A., Harris N.,
 RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Seale S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of Drosophila melanogaster genome";
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 [5]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AB003464; AAF47246.2; -.
 DR FlyBase; FBgn035041; CG13594.
 SQ SEQUENCE 478 AA; 49256 MW; 3C56261A6EB8F8A CRC64;
 Query Match 1.1%; Score 7; DB 5; Length 478;
 Best Local Similarity 100.0%; Pred. No. 5.7e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 187 VYLNASY 193
 Db 19 VYLNASY 25

```

RESULT 28
Q9S323      PRELIMINARY;      PRT;      479 AA.
AC  Q9S323;
DT  01-MAY-2000 (TrEMBLrel. 13, Created)
DT  01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE  01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE  Acyl-CoA reductase LuxC.
GN  LuxC.
OS  Vibrio fischeri.
OC  Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC  Vibrionaceae; Vibrio.
OX  NCBI_TaxID=668;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=MJ-1;
RA  "Vibrio fischeri lux operon SalI digest.";
RT  Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RL  EMBL; AF170104; AAD48475.1; -.
DR  GO; GO:0019109; F:acyl-CoA reductase activity; IEA.
DR  GO; GO:0008218; P:bioluminescence; IEA.
DR  InterPro; IPR008670; LuxC.
DR  Pfam; PF05893; LuxC; 1.
SQ  SEQUENCE 479 AA; 54578 MW; 10FD9567E5026340 CRC64;

Query Match      1.1%; Score 7; DB 2; Length 479;
Best Local Similarity 100.0%; Pred. No. 5.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  117 SVHLNG 123
DB  146 SVHLNG 152

RESULT 29
Q9LZG8      PRELIMINARY;      PRT;      479 AA.
AC  Q9LZG8;
DT  01-OCT-2000 (TrEMBLrel. 15, Created)
DT  01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT  01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE  Transporter-like protein.
GN  T8A8_80.
OS  Arabidopsis thaliana (Mouse-ear cress).
OC  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC  eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX  NCBI_TaxID=3702;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Parnell B., Boutry M., Goffeau A., Mewes H.W., Rudd S., Lemcke K.,
RA  Mayer K.F.X., Quelet F., Salanoubat M.;
RL  Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN  [2]
RP  SEQUENCE FROM N.A.
RA  EU Arabidopsis sequencing project;
RA  Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC  -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR  EMBL; AL162691; CAB83151.1; -.
DR  PIR; T47415; T47415.
DR  GO; GO:0016021; C:integral to membrane; IEA.
DR  GO; GO:0005215; F:transporter activity; IEA.
DR  GO; GO:0006810; P:transport; IEA.
DR  InterPro; IPR007114; MFS.
DR  InterPro; IPR005828; Sub transporter.
DR  Pfam; PF00083; sugar tr; 1.
DR  PROSITE; PSS0850; MFS; 1.
DR  Transmembrane.
SQ  SEQUENCE 479 AA; 52806 MW; 8344F89F8B3E3CA CRC64;

Query Match      1.1%; Score 7; DB 10; Length 479;

```

```

Best Local Similarity 100.0%; Pred. No. 5.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  518 LFGSTS 524
DB  117 LFGSTS 123

RESULT 30
Q56817      PRELIMINARY;      PRT;      480 AA.
AC  Q56817;
DT  01-NOV-1996 (TrEMBLrel. 01, Created)
DT  01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT  01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE  Fatty acid reductase.
GN  LuxC.
OS  Photobacterium luminescens (Xenorhabdus luminescens).
OC  Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC  Enterobacteriaceae; Photobacterium.
OX  NCBI_TaxID=29488;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Hb;
RX  MEDLINE=9235513; PubMed=1644764;
RA  "Multiple repetitive elements organization of the lux operons of
RT  luminescent terrestrial bacteria.";
RT  J. Bacteriol. 174:5371-5381 (1992).
DR  EMBL; M90093; AAA27617.1; -.
DR  GO; GO:0019109; F:acyl-CoA reductase activity; IEA.
DR  GO; GO:0008218; P:bioluminescence; IEA.
DR  InterPro; IPR008670; LuxC.
DR  Pfam; PF05893; LuxC; 1.
SQ  SEQUENCE 480 AA; 54636 MW; AD42A2BF091B2B3 CRC64;

Query Match      1.1%; Score 7; DB 2; Length 480;
Best Local Similarity 100.0%; Pred. No. 5.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  117 SVHLNG 123
DB  147 SVHLNG 153

RESULT 31
Q8SY11      PRELIMINARY;      PRT;      480 AA.
AC  Q8SY11;
DT  01-JUN-2002 (TrEMBLrel. 21, Created)
DT  01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT  01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE  RE16208p.
GN  CG13594.
OS  Drosophila melanogaster (Fruit fly).
OC  Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC  Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC  Ephydroidea; Drosophilidae; Drosophila.
OX  NCBI_TaxID=7227;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=Berkley;
RA  Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA  Champe M., Chavez C., Dorsett V., Drensek D., Farfan D., Frise B.,
RA  George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA  Miranda A., Mungall C.J., Nunoo J., Pacleb J., Pargass V., Park S.,
RA  Patel S., Phouenavong S., Wan K., Yu C., Lewis S.B., Rubin G.M.,
RA  Celinker S.;
RL  Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AY075460; AAL68273.1; -.
DR  FlyBase; FBgn0035041; CG13594.
SQ  SEQUENCE 480 AA; 49470 MW; 052E1B3317EDD700 CRC64;

```

Query Match 1.1%; Score 7; DB 5; Length 480;
 Best Local Similarity 100.0%; Pred. No. 5.8e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 187 YVUNASY 193
 |||||
 21 YVUNASY 27

RESULT 32

Q9H6A7 PRELIMINARY; PRT; 483 AA.
 AC Q9H6A7;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
 DE Hypothetical protein FLJ22429.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 ON NCBI_TaxID=9606;
 RX [1]

RP SEQUENCE FROM N.A.
 RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 RA Okitani R., Ota T., Suzuki Y., Odayashi M., Nishi T., Shibahara T.,
 RA Tanaka T., Nakamura Y., Isegaki T., Sugano S.;
 RT "NEDO human cDNA sequencing project.";
 RU Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]

RP SEQUENCE FROM N.A.
 RC TISSUE=Ovary;

RA Strauberg R.;

RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AK026082; BAB15355.1; -.

DR EMBL; BC012837; AAH12837.1; -.

KW Hypothetical protein.

SQ SEQUENCE 483 AA; 53266 MW; C16B129E6005BAAF CRC64;

Query Match

1.1%; Score 7; DB 4; Length 483;
 Best Local Similarity 100.0%; Pred. No. 5.8e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 487 VRTSSSG 493
 |||||
 53 VRTSSSG 59

RESULT 33

Q9SXY1 PRELIMINARY; PRT; 487 AA.

AC Q9SXY1;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)

DE FAS2 (Fragment).

GN FAS2.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.

ON NCBI_TaxID=3702;

RX [1]

RP SEQUENCE FROM N.A.

RC STRAIN=COLUMBIA;

RA Arai T., Kaya H.;

RT "FASCIAT2, open reading frame.";

RU Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Columbia;

RX MEDLINE=98290546; PubMed=9628582;

RA Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,
 Tabata S.;

RT "Structural analysis of Arabidopsis thaliana chromosome 5. IV.
 RT Sequence features of the regions of 1,456,315 bp covered by nineteen
 RT physically assigned pl and TAC clones.";

RL DNA Res. 5:41-54(1998).

DR EMBL; AB027230; BAB77766.1; -.

DR EMBL; AB010076; BAB1430.1; -.

DR InterPro; IPR001680; WD40.

DR Pfam; PF00400; WD40; 5.

DR SMART; SM00320; WD40; 5.

DR PROSITE; PS00678; WD_REPEATS_1; 1.

DR PROSITE; PS00682; WD_REPEATS_2; 3.

DR PROSITE; PS50294; WD_REPEATS_REGION; 1.

KW Repeat; WD repeat.

FT NON TER 487

SQ SEQUENCE 487 AA; 54106 MW; 06CB34C29E1F3191 CRC64;

Query Match

1.1%; Score 7; DB 10; Length 487;
 Best Local Similarity 100.0%; Pred. No. 5.8e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 95 VASLSSD 101
 |||||
 169 VASLSSD 175

RESULT 34

Q8ZJT2 PRELIMINARY; PRT; 488 AA.

AC Q8ZJT2;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)

DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)

DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)

DE Hypothetical protein YPO0004.

GN YPO0004 OR Y0004.

OS Yersinia pestis.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Yersinia.

ON NCBI_TaxID=632;

RX [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CO-92 / Biovar Orientalis;

RE MEDLINE=21470413; PubMed=11586360;

RA Parkhill J., Wren B.W., Thomson N.R., Titchall R.W., Holden M.T.G.,

RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mangal K.L.,

RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,

RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,

RA Feltham S., Hamlin N., Holtroyd S., Jags K., Karlyshev A.V.,

RA Leather S., Moule S., Oyelson P.C.F., Quail M., Rutherford K.,

RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;

RT "Genome sequence of Yersinia pestis, the causative agent of plague.";

RU Nature 413:523-527(2001).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=KIM5 / Biovar Mediaevalis;

RE MEDLINE=22137863; PubMed=12142430;

RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Lies P.,

RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,

RA Fetherston J.D., Lindner L.E., Brubaker R.R., Plano G.V.,

RA Straley S.C., McDonough K.A., Niles M.L., Matson J.S., Blattner F.R.,

RA Perry R.D.;

RT "Genome sequence of Yersinia pestis KIM.";

RU J. Bacteriol. 184:4601-4611(2002).

DR EMBL; AJ414141; CAC88871.1; -.

DR EMBL; AE013601; AAM83600.1; -.

DR PIR; AE0001; AE0001.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 488 AA; 56081 MW; CD17471A7A85D82D CRC64;

QY 369 SLEEDSP 375

DB 80 SLEEDSP 86

RESULT 35

OBIMG2 PRELIMINARY; PRT; 490 AA.

AC OBIMG2; 01-JUN-2003 (T-EMBLrel. 24, Created)
 DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE Stage V sporulation protein AF.
 GN SPOVAF OR BA4286.
 OS Bacillus anthracis (strain Ames).
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 NCBI_TaxID=198094;
 OX [1]

SEQUENCE FROM N.A.

RX MEDLINE=22608414; PubMed=12721629;
 RA Read T.D., Peterson S.N., Tourasse N., Baillie L.W., Paulsen I.T.,
 Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,
 Holtapple E.K., Okstad O.A., Helgason E., Ristic J., Wu M.,
 Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.,
 DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,
 Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,
 Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,
 Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,
 Hazen A., Cline R., Redmond C., Thwaite J.E., White O., Salzberg S.L.,
 Thomson B., Friedlander A.M., Koehler T.M., Hanna P.C., Kolsto A.-B.,
 Fraser C.M.;
 RA "The genome sequence of Bacillus anthracis Ames and comparison to
 RT closely related bacteria."
 RL Nature 423:81-86(2003).
 DR EMBL; AB017037; AAP28006.1; --
 DR TIGR; BA4286; --
 DR GO; GO:0016021; C:integral to membrane; IEA.
 DR GO; GO:0009847; P:spore germination; IEA.
 DR InterPro; IPR004995; GerA.
 DR Pfam; PF03323; GerA; 1.
 KW Complete proteome.
 SQ SEQUENCE 490 AA; 55064 MW; 02EBF557EB86009C CRC64;

Query Match 1.1%; Score 7; DB 16; Length 490;
 Best Local Similarity 100.0%; Pred. No. 5.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 377 VQALSGI 383
 DB 104 VQALSGI 110

RESULT 36

OBKXDB PRELIMINARY; PRT; 492 AA.

AC OBKXDB; 01-OCT-2000 (T-EMBLrel. 15, Created)
 DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE Hypothetical protein.
 GN H0115.
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 NCBI_TaxID=83334;
 OX [1]

SEQUENCE FROM N.A.

RP STRAIN=O157:H7;
 RX MEDLINE=20198760; PubMed=10734605;
 RA Makino K., Yokoyama K., Kubota Y., Yutsudo C.H., Kimura S.,
 RA Kurokawa K., Ishii K., Hattori M., Tatsuno I., Abe H., Iida T.,
 RA Yamamoto K., Ohnishi M., Hayashi T., Yasunaga T., Honda T.,
 RA Sasakawa C., Shinagawa H.;
 RT "Complete nucleotide sequence of the prophage VT2-Sakai carrying the

RT verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
 derived from the Sakai outbreak.";
 RL Genes Genet. Syst. 74:227-239(1999).
 DR EMBL; AF000422; BAA94143.1; --
 DR InterPro; IPR005181; DUF303.
 DR Pfam; PF03629; DUF303; 1.
 KW Hypothetical protein.

SQ SEQUENCE 492 AA; 52888 MW; 42BEFA7F2274B0DC CRC64;

Query Match 1.1%; Score 7; DB 2; Length 492;
 Best Local Similarity 100.0%; Pred. No. 5.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 417 LHGFSSS 423
 DB 475 LHGFSSS 481

RESULT 37

OBZYH4 PRELIMINARY; PRT; 494 AA.

AC OBZYH4; 01-MAR-2002 (T-EMBLrel. 20, Created)
 DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE Conserved protein (Possible type II secretion, gsp).
 GN PAE0773.
 OS Pyrobaculum aerophilum.
 OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;
 OC Thermoproteaceae; Pyrobaculum.
 NCBI_TaxID=13773;
 OX [1]

RX SEQUENCE FROM N.A.
 RC STRAIN=IM2 / ATCC 51768 / DSM 7523;
 RX MEDLINE=2164397; PubMed=11792869;
 RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,
 RA Miller J.H.;
 RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum
 aerophilum."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).
 DR EMBL; AE009784; AAL63019.1; --
 DR GO; GO:0005622; C:intracellular; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0008170; F:N-methyltransferase activity; IEA.
 DR GO; GO:0006306; F:DNA methylation; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR001482; GSP11_E.
 DR InterPro; IPR002052; N6_Mtase.
 DR Pfam; PF00437; GSP11_E; 1.
 DR Pfam; PF000739; GSP11_E; 1.
 DR PROSITE; PS00092; N6_MTASE; 1.

KW Complete proteome.
 SQ SEQUENCE 494 AA; 56443 MW; D70B2A55341B3B41 CRC64;

Query Match 1.1%; Score 7; DB 17; Length 494;
 Best Local Similarity 100.0%; Pred. No. 5.9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 642 SREELGK 648
 DB 69 SREELGK 75

RESULT 38

OBW0U3 PRELIMINARY; PRT; 495 AA.

AC OBW0U3; 01-MAY-2000 (T-EMBLrel. 13, Created)
 DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
 DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
 DE CG7042 protein.
 GN CG7042.

OS *Drosophila melanogaster* (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NC NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer V.G., Champe M., Pfeiffer B.D.,
 RA Wen K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abriil J.F., Agbayani A., An H.-U., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Paulis B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,
 RA Foeller C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Gloddek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spter E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirbas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Fafian D.,
 RA Ferriere S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nuno J.,
 RA Pacle J., Parag V., Park S., Patel S., Pfeiffer B.,
 RA Pounenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svirbas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC Hiera S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Klinkner J.S., Prochuk S.E., Smith C.D.,
 RA Thuy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
 RA Clump M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Krommiller B., Marshall B., Milburn G., Richer J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutnack F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;

RT "Annotation of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RC FlyBase;
 RA Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RX EMBL; AF003467; AF47341.2; -;
 DR FlyBase; FBgn0035105; CG7042.
 DR GO; GO:0008138; F:protein tyrosine/serine/threonine phosphata...; IEA.
 DR GO; GO:0006470; P:protein amino acid dephosphorylation; IEA.
 DR InterPro; IPR001406; Pseudocatalase.
 DR InterPro; IPR001406; Pseudocatalase.
 DR InterPro; IPR000387; TYR_phosphatase.
 DR Pfam; PF00782; DSPC; 1.
 DR SMART; SM00195; DSPC; 1.
 DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE; PS00386; TYR_PHOSPHATASE_2; 1.
 DR PROSITE; PS00504; TYR_PHOSPHATASE_DUAL; 1.
 SQ SEQUENCE 495 AA; 56479 MW; 9A0BDF2A42069C3 CRC64;
 Query Match 1.1%; Score 7; DB 5; Length 495;
 Best Local Similarity 100.0%; Pred. No. 5; 9e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 238 GCYLVHC 244
 Db 438 GCYLVHC 444
 RESULT 39
 ID 076107 PRELIMINARY; PRT; 499 AA.
 AC 076107;
 DT 01-NOV-1998 (TEMBLrel. 08, Created)
 DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
 DE Cytochrome P450.
 GN CYP6B1.
 OS *Culex quinquefasciatus* (Southern house mosquito).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Culex.
 OC NCBI_TaxID=7176;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JPAL-per; TISSUE=gut;
 RX MEDLINE=98194786; PubMed=9535163;
 RA Kasai S., Shono T., Yamakawa M.;
 RT "Molecular cloning and nucleotide sequence of a cytochrome P450 cDNA
 from a pyrethroid-resistant mosquito, *Culex quinquefasciatus* Say.";
 RL Insect Mol. Biol. 7:185-190(1998).
 CC 1- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
 DR EMBL; AB001323; BAA28946.1; -;
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0004497; F:monooxygenase activity; IEA.
 DR GO; GO:0008170; F:N-methyltransferase activity; IEA.
 DR GO; GO:0006306; P:DNA methylation; IEA.
 DR GO; GO:0006118; P:electron transport; IEA.
 DR InterPro; IPR001128; Cytochrome_P450.
 DR InterPro; IPR002052; N6_Mtase.
 DR Pfam; PF00067; P450; 1.
 DR PRINTS; PR00385; P450.
 DR PROSITE; PS00086; CYTOCHROME_P450; 1.
 DR PROSITE; PS00092; N6_MTASE; 1.
 KW Heme; Monooxygenase; Oxidoreductase.
 SQ SEQUENCE 499 AA; 57486 MW; A1C42A04E9ECF238 CRC64;
 Query Match 1.1%; Score 7; DB 5; Length 499;
 Best Local Similarity 100.0%; Pred. No. 6e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 544 LAPOTST 550
|||||
Db 455 LAPOTST 461

RESULT 40

ID Q9N3R4 PRELIMINARY; PRT; 502 AA.
AC Q9N3R4;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein.
GN Y47G6A.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Minx P., Graves T., Hawrysko C.;
RT "The sequence of C. elegans cosmid Y47G6A.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DE EMBL: AC024791; AAF60654.2; -.
DR WormBep; Y47G6A.1; CE27258.
DR InterPro; IPR000990; Inmexin.
DR Pfam; PF00876; Inmexin; 1.
DR PRINTS; PR01262; INNEXIN.
KW Hypothetical protein.
SQ SEQUENCE 502 AA; 57540 MW; 29178A4D71580D0D CRC64;

Query Match 1.1%; Score 7; DB 5; Length 502;
Best Local Similarity 100.0%; Pred. No. 6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 LBSGTEK 25
|||||
Db 40 LBSGTEK 46

RESULT 41

ID O806Y8 PRELIMINARY; PRT; 505 AA.
AC O806Y8;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE RNA-dependent RNA polymerase (Fragment).
GN RDRP.
OS Cucumber yellows virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Closteroviridae;
OC Crinivirus.
OX NCBI_TaxID=32618;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22541874; PubMed=12655104;

RA Hartono S., Natsunaki T., Genda Y., Okuda S.;
RT "Nucleotide sequence and genome organization of Cucumber yellows
virus, a member of the genus Crinivirus.";
RL J. Gen. Virol. 84:1007-1012(2003).
RN EMBL; AB085613; BAC66371.1; -.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0003968; F:RNA-directed RNA polymerase activity; IEA.
DR GO; GO:0006350; P:transcription; IEA.
DR GO; GO:0019079; P:viral genome replication; IEA.
DR InterPro; IPR001788; RNA_dep_RNApol2.
DR InterPro; IPR007095; RNA_pol_DS_Ps.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR Pfam; PF00978; RNA_dep_RNApol2; 1.
KW RNA-directed RNA polymerase.
FT NON TER 1
SQ SEQUENCE 505 AA; 58437 MW; 877D9E2286B6DB5 CRC64;

Query Match 1.1%; Score 7; DB 12; Length 505;
Best Local Similarity 100.0%; Pred. No. 6e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 395 KTKRSFS 401
|||||
Db 166 KTKRSFS 172

RESULT 42

ID Q7UH12 PRELIMINARY; PRT; 508 AA.
AC Q7UH12;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE P0P1-Arpaase beta chain.
GN ATPD OR RB4906.
OS Rhodospirillum rubrum.
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC Planctomycetaceae; Pirellula.
OX NCBI_TaxID=117;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1;
RX MEDLINE=22735913; PubMed=12835416;
RA Glockner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Borzym K., Hellmann K., Rabus R.,
RA Schlesner H., Amann R., Reinhardt R.;
RT "Complete genome sequence of the marine planctomycete Pirellula sp.
strain 1.";
RT Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).
DR EMBL; BX294141; CAD78167.1; -.
KW Complete proteome.
SQ SEQUENCE 508 AA; 55235 MW; F4E4A544A48BCFB9 CRC64;

Query Match 1.1%; Score 7; DB 16; Length 508;
Best Local Similarity 100.0%; Pred. No. 6.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 63 VLITELI 69
|||||
Db 210 VLITELI 216

RESULT 43

ID O50188 PRELIMINARY; PRT; 509 AA.
AC O50188; O50187;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Probable phosphoprotein phosphatase.
GN ppp.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 [1]
 RP SEQUENCE FROM N.A.
 RA Cole S.T.; (AUG-1997) to the EMBL/GenBank/DBJ databases.
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93189700; PubMed=8446027;
 RA Eigemeier K., Honore N., Woods S.A., Caudron B., Cole S.T.;
 RT "Use of an ordered cosmid library to deduce the genomic organization
 of Mycobacterium leprae.";
 RL Mol. Microbiol. 7:197-206(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97124199; PubMed=8969512;
 RA Feihl H., De Rossi E., Salazar L., Cantoni R., Labo M., Riccardi G.,
 RA Takiff H.E., Eigemeier K., Bergh S., Cole S.T.;
 RT "Gene arrangement and organization in a approximately 76 kb fragment
 encompassing the orfC region of the chromosome of Mycobacterium
 leprae.";
 RL Microbiol. 142:3147-3161(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96310367; PubMed=8733228;
 RA Salazar L., Feihl H., De Rossi E., Riccardi G., Rios C., Cole S.T.,
 RA Takiff H.E.;
 RT "Organization of the origins of replication of the chromosomes of
 Mycobacterium smegmatis, Mycobacterium leprae and Mycobacterium
 tuberculosis and isolation of a functional origin from M. smegmatis.";
 RL Mol. Microbiol. 20:283-293(1996).
 RN [5]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96194983; PubMed=8622949;
 RA Feihl H., Vincent V., Cole S.T.;
 RT "Homing events in the gyrA gene of some mycobacteria.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:3410-3415(1996).
 DR EMBL; Z70722; CAA94721.1; -
 DR PIR; T10013; T10013.
 DR GO; GO:0003824; F:catalytic activity; IEA.
 DR InterPro; IPR001932; PP2C-like.
 DR InterPro; IPR003643; Pto_Pase.
 DR Pfam; PF00481; PP2C; 1.
 DR Prodom; PD006823; Pto_Pase; 1.
 DR SMART; SM00332; PP2C_1.
 DR SMART; SM00331; PP2C_SIG; 1.
 DR SQUENCE 509 AA; 53661 MW; 7D8DC9691186DF9C CRC64;
 SQ

Query Match 1.1%; Score 7; DB 2; Length 509;
 Best Local Similarity 100.0%; Pred. No. 6.1e+02; Indels 0; Gaps 0;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 489 TSSSGTA 495
 |||||
 Db 460 TSSSGTA 466

RESULT 44
 O9CDES PRELIMINARY; PRT; 509 AA.
 AC O9CDES;
 DT 01-JUN-2001 (TEMBLrel. 17, Created)
 DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
 DE Hypothetical protein ML0020.
 GN ML0020.
 OS Mycobacterium leprae.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TW;

RX MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Eigemeier K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Fellgett T., Fraser A., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagsels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrall B.G.;
 RT "Massive gene decay in the leprosy bacillus.";
 RL Nature 409:1007-1011(2001).
 DR EMBL; AL583917; CAC29528.1; -
 DR PIR; D86911; D86911.
 DR Leproma; ML0020; -
 DR GO; GO:0003824; F:catalytic activity; IEA.
 DR InterPro; IPR001932; PP2C-like.
 DR InterPro; IPR003643; Pto_Pase.
 DR Pfam; PF00481; PP2C; 1.
 DR Prodom; PD006823; Pto_Pase; 1.
 DR SMART; SM00332; PP2C_1.
 DR SMART; SM00331; PP2C_SIG; 1.
 DR KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 509 AA; 53576 MW; 6D0B3DF76D9D1ED9 CRC64;
 S

Query Match 1.1%; Score 7; DB 16; Length 509;
 Best Local Similarity 100.0%; Pred. No. 6.1e+02; Indels 0; Gaps 0;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 489 TSSSGTA 495
 |||||
 Db 460 TSSSGTA 466

RESULT 45
 O8YGD2 PRELIMINARY; PRT; 511 AA.
 AC O8YGD2;
 DT 01-MAR-2002 (TEMBLrel. 20, Created)
 DT 01-MAR-2002 (TEMBLrel. 20, Last sequence update)
 DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
 DE Hypothetical protein BME11227.
 GN BME11227.
 OS Brucella melitensis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Brucella.
 OX NCBI_TaxID=29459;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=16M / ATCC 23456 / Biotype 1;
 RX MEDLINE=20020109; PubMed=11756588;
 RA DelVecchio V.G., Kapatral V., Redkar R.J., Patra G., Mijer C., Los T.,
 RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
 RA Jablonksi L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,
 RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
 RA Haselkorn R., Kyrides N., Overbeek R.;
 RT "The genome sequence of the facultative intracellular pathogen
 Brucella melitensis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
 DR EMBL; AE009562; AL52408.1; -
 DR PIR; AE3405; AE3405.
 DR GO; GO:0008080; F:N-acetyltransferase activity; IEA.
 DR InterPro; IPR000182; GCN5acetyl_trans.
 DR Pfam; PF00583; Acetyltransf; 1.
 DE Hypothetical protein; Complete proteome.
 SQ SEQUENCE 511 AA; 57832 MW; 729C9855EBB414F8 CRC64;
 S

Query Match 1.1%; Score 7; DB 16; Length 511;
 Best Local Similarity 100.0%; Pred. No. 6.1e+02; Indels 0; Gaps 0;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 73 AKHKVDI 79
 |||||

Db 34 AKHKVDI 40

RESULT 46

ID 08G1J0 PRELIMINARY; PRT; 511 AA.

AC 08G1J0; (1-MAR-2003 (TREMBlrel. 23, Created))

DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)

DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

GN Acetyltransferase, GNAT family.

OS BR0725.

OC Brucella suis.

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

OC Brucellaceae; Brucella.

OX NCBI_TaxID=29461;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=1330 / Biovar 1;

RE MEDLINE=22247741; PubMed=12271122;

RA Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,

RA Read T.D., Dodson R.D., Umayam L., Brinkac L.M., Beanan M.J.,

RA Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,

RA Nelson W.C., Ayodeji B., Kirati M., Shetty J., Malek J., Van Aken S.B.,

RA Riedmuller S., Tettelin H., Gill S.R., White O., Salzberg S.L.,

RA Hoover D.L., Iandier L.E., Halling S.M., Boyle S.M., Fraser C.M.,

RT "The Brucella suis genome reveals fundamental similarities between

animal and plant pathogens and symbionts."

RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).

DR EMBL; AE014378; AAN29654.1; -.

DR TIGR; BR0725; -.

DR GO; GO:0008080; P:N-acetyltransferase activity; IEA.

DR GO; GO:0016740; F:transferase activity; IEA.

DR InterPro; IPR000182; GCNacetyl_trans.

DR Pfam; PF00583; Acetyltransf. 1.

DR TRANSFAC; Complete proteome.

SQ SEQUENCE 511 AA; 57628 MW; 411004CA35883AAE CRC64;

Query Match 1.1%; Score 7; DB 16; Length 511;

Best Local Similarity 100.0%; Pred. No. 6.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 73 AKHKVDI 79

Db 34 AKHKVDI 40

RESULT 47

ID P77815 PRELIMINARY; PRT; 515 AA.

AC P77815; (1-FEB-1997 (TREMBlrel. 02, Created))

DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)

DE 01-JUN-2003 (TREMBlrel. 24, Last annotation update)

GN 3-ketosteroid-1-dehydrogenase (EC 1.3.99.4).

OS KSDP.

OC Nocardioideae simplex (Arthrobacter simplex).

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Propionibacteriineae; Nocardioideae; Pimelobacter.

OX NCBI_TaxID=2045;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=IFO12096;

RE MEDLINE=95319331; PubMed=7596291;

RA Molnar I., Choi K., Yamashita M., Murooka Y.,

RT "Molecular cloning, expression in Streptomyces lividans, and analysis

of a gene cluster from Arthrobacter simplex encoding 3-

ketosteroid-DELTA.1-dehydrogenase, 3-ketosteroid-DELTA.5-isomerase

RT and a hypothetical regulatory protein."

RL Mol. Microbiol. 15:895-905(1995).

DR EMBL; D37969; BAA07186.1; -.

DR PIR; S61889; S61889.

DR GO; GO:0016020; C:membrane; IEA.

DR GO; GO:0016491; F:oxidoreductase activity; IEA.

DR GO; GO:0005215; F:transporter activity; IEA.

DR GO; GO:0006810; P:transport; IEA.

DR InterPro; IPR005829; Sug_transporter.

DR PROSITE; PS00216; SUGAR_TRANSPORT_1; 1.

KW Oxidoreductase.

SQ SEQUENCE 515 AA; 54330 MW; 78CFECA65AB022DB CRC64;

Query Match 1.1%; Score 7; DB 2; Length 515;

Best Local Similarity 100.0%; Pred. No. 6.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 424 EDALERY 430

Db 71 EDALERY 77

RESULT 48

ID 082NP2 PRELIMINARY; PRT; 516 AA.

AC 082NP2; (1-JUN-2003 (TREMBlrel. 24, Created))

DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)

DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE Putative acyl-CoA synthetase, long-chain fatty acid:CoA ligase.

GN FAD6 OR SAV1259.

OS Streptomyces avermitilis.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Streptomyetaceae; Streptomyetaceae; Streptomyces.

OX NCBI_TaxID=33903;

RN (1)

RP SEQUENCE FROM N.A.

RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;

RE MEDLINE=22477403; PubMed=11572948;

RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,

RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,

RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.,

RT "Genome sequence of an industrial microorganism Streptomyces

RT avermitilis: deducing the ability of producing secondary

metabolites."

RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).

RN (2)

RP SEQUENCE FROM N.A.

RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;

RE MEDLINE=22608306; PubMed=12692562;

RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,

RA Sakaki Y., Hattori M., Omura S.,

RT "Complete genome sequence and comparative analysis of the industrial

microorganism Streptomyces avermitilis."

RL Nat. Biotechnol. 21:526-531(2003).

DR EMBL; AP005026; BAC68969.1; -.

DR GO; GO:0016874; F:Ligase activity; IEA.

DR GO; GO:0008152; P:metabolism; IEA.

DR InterPro; IPR000873; AMP-bind.

DR Pfam; PF00501; AMP-binding; 1.

DR PROSITE; PS00455; AMP BINDING; 1.

KW Ligase; Complete proteome.

SQ SEQUENCE 516 AA; 56513 MW; 6255B86728639A6A CRC64;

Query Match 1.1%; Score 7; DB 16; Length 516;

Best Local Similarity 100.0%; Pred. No. 6.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 225 KSVDFIE 231

Db 483 KSVDFIE 489

RESULT 49

ID Q93N44 PRELIMINARY; PRT; 519 AA.

AC Q93N44; (01-DEC-2001 (TREMBlrel. 19, Created))

DT 01-DEC-2001 (TREMBlrel. 19, Created)

Search completed: June 21, 2004, 13:28:28
 Job time : 52 secs

DT 01-DEC-2001 (TrEMBLrel. 19, last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, last annotation update)
 DE Sulfolobus ferredoxin (Hypothetical protein).
 GN CBU0699.
 OS Coxiella burnetii.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
 OC Coxiellaceae; Coxiella.
 OC NCBI_TaxID=777;
 OX [1]
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Hoover T.A., Vodkin M.H., Williams J.C., Culp D.W., Thompson H.A.;
 RT "A chromosomal DNA deletion explains the phenotype of the Coxiella
 burnetii phase II variant."
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Nine Mile phase I / RSA 493;
 RX MEDLINE=22608657; PubMed=12704232;
 RA Seeshadri R., Paulsen I.T., Eisen J.A., Read T.D., Nelson K.E.,
 RA Nelson W.C., Ward N.L., Tettelin H., Davidson T.M., Beaman M.J.,
 RA DeBoy R.T., Daugherty S.C., Brinkac L.M., Madupu R., Dodson R.J.,
 RA Khouri H.M., Lee K.H., Carty H.A., Scanlan D., Heinzen R.A.,
 RA Thompson H.A., Samuel J.E., Fraser C.M., Heidelberg J.F.;
 RT "Complete genome sequence of the Q-fever pathogen, Coxiella
 burnetii."
 RT Proc. Natl. Acad. Sci. U.S.A. 100:5455-5460(2003).
 RL EMBL: AF387640; AAK71278.1; -.
 DR EMBL: AEO16862; AAO90243.1; -.
 DR TIGR: CB00699; -.
 DR GO: GO:0016740; F:transferase activity; IEA.
 DR Hypothetical protein; Transferase; Complete proteome.
 SQ SEQUENCE 519 AA; 61647 MW; 9D6E70B81AC690C4 CRC64;

Query Match
 Best Local Similarity 100.0%; Score 7; DB 16; Length 519;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 43 LEANIN 49
 |||||
 DB 36 LEANIN 42

RESULT 50
 O8N3D9
 ID O8N3D9 PRELIMINARY; PRT; 523 AA.
 AC O8N3D9;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, last annotation update)
 DE Hypothetical protein (Fragment).
 GN DKFZP762F135.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Melanoma;
 RA Bloembergen H., Boecher M., Brandt P., Mewes H.W., Weil B., Wiemann S.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL834428; CAD39089.1; -.
 KW Hypothetical protein.
 FT NON_TER 1
 SQ SEQUENCE 523 AA; 54448 MW; 0B58F308E13320F8 CRC64;

Query Match
 Best Local Similarity 100.0%; Score 7; DB 4; Length 523;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 356 HPASVPS 362
 |||||
 DB 341 HPASVPS 347

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 13:21:13 ; Search time 58 Seconds
(without alignments)
3239.553 Million cell updates/sec

Title: US-10-029-345A-109

Perfect score: 665

Sequence: 1 MAHEMIGTIVTERLVALL.....LKVSGSGSFGSGMEIIEVS 665

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 0

Total number of hits satisfying chosen parameters: 162285

Minimum DB seq length: 473

Maximum DB seq length: 200000000

Post-Processing: Listing first 75 summaries

Database : A_Geneseq_29yand4:*

1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	665	100.0	665	5	ABR52381
2	665	100.0	665	5	ABR52407
3	472	71.0	517	5	AAU79159
4	472	71.0	665	4	AAE04834
5	472	71.0	665	4	AAU09016
6	472	71.0	665	5	AAU79156
7	472	71.0	665	5	AAU09946
8	472	71.0	665	5	AAU75789
9	472	71.0	665	5	ABR57946
10	472	71.0	665	5	AAU79929
11	472	71.0	665	5	ABR57291
12	472	71.0	665	6	ADA54744
13	472	68.0	665	5	AAU79161
14	472	68.0	665	5	AAU79162
15	350	52.6	665	5	ABR52352
16	281	42.3	664	5	ABR52424
17	266	40.0	666	4	AAAB20325
18	258	38.8	672	4	AAAB20325
19	64	9.6	660	5	ABR52385
20	20	3.0	579	7	ABR08458
21	20	3.0	625	5	ABR52382
22	20	3.0	625	5	ABR52350
23	20	3.0	625	6	ABG73440
24	20	3.0	663	2	AAW29150
25	20	3.0	663	5	ABR52351

26	13	2.0	476	4	ABR63527	Abb63527 Drosophila
27	12	1.8	436	4	ABG00724	Abg00724 Novel hum
28	8	1.2	876	7	ADA44807	Ada44807 CD4/CCR C
29	8	1.2	474	3	AAV59170	AAV59170 CD4-Ig fu
30	8	1.2	481	1	AAV59101	AAV59101 Genetic C
31	8	1.2	481	3	AAV51081	AAV51081 CD4-IgM f
32	8	1.2	481	3	AAV51081	AAV51081 Human fus
33	8	1.2	481	3	AAV59171	AAV59171 CD4-Ig fu
34	8	1.2	507	3	AAV59171	AAV59171 CD4-Ig fu
35	8	1.2	519	2	AAV59171	AAV59171 CD4-Ig fu
36	8	1.2	524	1	AAV59171	AAV59171 CD4-Ig fu
37	8	1.2	530	2	AAV59171	AAV59171 CD4-Ig fu
38	8	1.2	530	2	AAV59171	AAV59171 CD4-Ig fu
39	8	1.2	530	2	AAV59171	AAV59171 CD4-Ig fu
40	8	1.2	530	2	AAV59171	AAV59171 CD4-Ig fu
41	8	1.2	530	2	AAV59171	AAV59171 CD4-Ig fu
42	8	1.2	530	2	AAV59171	AAV59171 CD4-Ig fu
43	8	1.2	532	2	AAV59171	AAV59171 CD4-Ig fu
44	8	1.2	532	2	AAV59171	AAV59171 CD4-Ig fu
45	8	1.2	532	2	AAV59171	AAV59171 CD4-Ig fu
46	8	1.2	532	2	AAV59171	AAV59171 CD4-Ig fu
47	8	1.2	532	2	AAV59171	AAV59171 CD4-Ig fu
48	8	1.2	534	2	AAV59171	AAV59171 CD4-Ig fu
49	8	1.2	549	2	AAV59171	AAV59171 CD4-Ig fu
50	8	1.2	557	2	AAV59171	AAV59171 CD4-Ig fu
51	8	1.2	574	2	AAV59171	AAV59171 CD4-Ig fu
52	8	1.2	575	2	AAV59171	AAV59171 CD4-Ig fu
53	8	1.2	575	2	AAV59171	AAV59171 CD4-Ig fu
54	8	1.2	575	2	AAV59171	AAV59171 CD4-Ig fu
55	8	1.2	575	2	AAV59171	AAV59171 CD4-Ig fu
56	8	1.2	575	2	AAV59171	AAV59171 CD4-Ig fu
57	8	1.2	575	2	AAV59171	AAV59171 CD4-Ig fu
58	8	1.2	590	6	AAV59171	AAV59171 CD4-Ig fu
59	8	1.2	616	3	AAV59171	AAV59171 CD4-Ig fu
60	8	1.2	616	3	AAV59171	AAV59171 CD4-Ig fu
61	8	1.2	616	3	AAV59171	AAV59171 CD4-Ig fu
62	8	1.2	616	3	AAV59171	AAV59171 CD4-Ig fu
63	8	1.2	631	1	AAV59171	AAV59171 CD4-Ig fu
64	8	1.2	631	1	AAV59171	AAV59171 CD4-Ig fu
65	8	1.2	631	1	AAV59171	AAV59171 CD4-Ig fu
66	8	1.2	631	1	AAV59171	AAV59171 CD4-Ig fu
67	8	1.2	631	1	AAV59171	AAV59171 CD4-Ig fu
68	8	1.2	631	1	AAV59171	AAV59171 CD4-Ig fu
69	8	1.2	631	1	AAV59171	AAV59171 CD4-Ig fu
70	8	1.2	631	1	AAV59171	AAV59171 CD4-Ig fu
71	8	1.2	631	1	AAV59171	AAV59171 CD4-Ig fu
72	8	1.2	631	1	AAV59171	AAV59171 CD4-Ig fu
73	8	1.2	631	1	AAV59171	AAV59171 CD4-Ig fu
74	8	1.2	631	1	AAV59171	AAV59171 CD4-Ig fu
75	8	1.2	631	1	AAV59171	AAV59171 CD4-Ig fu

ALIGNMENTS

RESULT 1
ABR52381
ID ABR52381 standard; protein; 665 AA.
XX
XX ABR52381;
DT 19-JUN-2003 (first entry)
DE Protein relating to the invention SEQ ID NO: 109.
KW antiproliferative; hepatotropic; nephrotropic; antiarthritic;
KW antiproliferative; cardiant; cytotoxic; gene therapy; liver disease;
KW proliferative disorder; renal failure; cardiovascular disorder;
KW immunological disorder; arthritis; psoriasis; congenital heart defect;
KW congestive heart failure; cancer; human; mouse; yeast; rat; fruitfly.
XX
XX Homo sapiens.
OS

PN WO200257460-A2.
 XX
 PD 25-JUL-2002.
 XX
 PF 20-DEC-2001; 2001WO-US050459.
 XX
 PR 20-DEC-2000; 2000US-0256868P.
 PR 30-MAR-2001; 2001US-0280186P.
 PR 01-MAY-2001; 2001US-0287735P.
 PR 05-JUN-2001; 2001US-0295848P.
 PR 25-JUN-2001; 2001US-0300465P.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Jackson DG, Feder J, Nelson T, Mintier G, Ramanathan C, Lee L;
 PI Siemers N, Bol D, Schieven G, Finger J, Todderud CG, Bassolino D;
 PI Krystek S, Mcatee P, Suchard S, Banas D;
 XX
 DR WPI; 2002-599721/64.
 DR N-PSDB; ACC60559.
 XX
 PT Novel polynucleotides encoding human phosphatase polypeptides useful in
 PT the prevention or treatment of e.g. proliferative and cardiovascular
 PT disorders.
 XX
 PS Claim 5; Fig 12; 801pp; English.
 XX
 CC The invention relates to a novel isolated nucleic acid comprising a
 CC polynucleotide having a nucleotide sequence selected from 40
 CC polynucleotides fully defined in the specification. The polynucleotide of
 CC the invention has antiproliferative, hepatotropic, nephrotropic,
 CC antiarthritic, antiposrotatic, cardiant, and cytostatic activity. The
 CC polynucleotide may have a use in gene therapy. A polynucleotide or
 CC polypeptide of the invention is useful for preventing, treating or
 CC ameliorating a medical condition, e.g. a proliferative disorder. They are
 CC also useful for treating e.g. liver disease, renal failure, immunological
 CC disorders including arthritis and psoriasis, cardiovascular disorders
 CC such as congenital heart defects and congestive heart failure, and
 CC cancer. A method of the invention is useful for diagnosing a pathological
 CC condition or susceptibility to a condition in a subject. The present
 CC sequence is used in the exemplification of the invention
 XX
 SQ Sequence 665 AA;
 Query Match 100.0%; Score 665; DB 5; Length 665;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 665; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 361 PSVPSVQSLBDSPLVQALSGHLHLSADRLBDSNKLKRSFSLDIKSVSYASMAASLHG 420
 QY 421 SSSDALBYKRPSTLLDGTNKLCOFSPVOELSEQPTSPDKKEASIRKLOTARPSDQ 480
 DB 421 SSSDALBYKRPSTLLDGTNKLCOFSPVOELSEQPTSPDKKEASIRKLOTARPSDQ 480
 QY 481 SKRLSVRTSSSGTQRSLSPILHRSGSVEDNVRHSFTGLSTSQOHLTKSAGLGLKGM 540
 DB 481 SKRLSVRTSSSGTQRSLSPILHRSGSVEDNVRHSFTGLSTSQOHLTKSAGLGLKGM 540
 QY 541 SDILAPQSTPSLTSSWYFATESSHFYASAIYGGASAYAVSCSQLPFCGQVYVRR 600
 DB 541 SDILAPQSTPSLTSSWYFATESSHFYASAIYGGASAYAVSCSQLPFCGQVYVRR 600
 QY 601 QKPSRPARSRMSWHEESPFKQFKRRSCOMEGESIMENRREBLGRVSGSSSTSGME 660
 DB 601 QKPSRPARSRMSWHEESPFKQFKRRSCOMEGESIMENRREBLGRVSGSSSTSGME 660
 QY 661 IIEVS 665
 DB 661 IIEVS 665
 RESULT 2
 ABR52407
 ID ABR52407 standard; protein; 665 AA.
 XX
 AC ABR52407;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Protein relating to the invention SEQ ID NO: 148.
 XX
 KW antiproliferative; hepatotropic; nephrotropic; antiarthritic;
 KW antiposrotatic; cardiant; cytostatic; gene therapy; liver disease;
 KW proliferative disorder; renal failure; cardiovascular disorder;
 KW immunological disorder; arthritis; psoriasis; congenital heart defect;
 KW congestive heart failure; cancer; human; mouse; yeast; rat; fruitfly.
 XX
 OS Homo sapiens.
 XX
 PN WO200257460-A2.
 XX
 PD 25-JUL-2002.
 XX
 PF 20-DEC-2001; 2001WO-US050459.
 XX
 PR 20-DEC-2000; 2000US-0256868P.
 PR 30-MAR-2001; 2001US-0280186P.
 PR 01-MAY-2001; 2001US-0287735P.
 PR 05-JUN-2001; 2001US-0295848P.
 PR 25-JUN-2001; 2001US-0300465P.
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 XX
 PI Jackson DG, Feder J, Nelson T, Mintier G, Ramanathan C, Lee L;
 PI Siemers N, Bol D, Schieven G, Finger J, Todderud CG, Bassolino D;
 PI Krystek S, Mcatee P, Suchard S, Banas D;
 XX
 DR WPI; 2002-599721/64.
 DR N-PSDB; ACC60572.
 XX
 PT Novel polynucleotides encoding human phosphatase polypeptides useful in
 PT the prevention or treatment of e.g. proliferative and cardiovascular
 PT disorders.
 XX
 PS Disclosure; Fig 19; 801pp; English.
 XX
 CC The invention relates to a novel isolated nucleic acid comprising a
 CC polynucleotide having a nucleotide sequence selected from 40
 CC polynucleotides fully defined in the specification. The polynucleotide of
 CC the invention has antiproliferative, hepatotropic, nephrotropic,

CC anti-infective, antiparasitic, cardiac, and cytoskeletal activity. The
CC polynucleotide may have a use in gene therapy. A polynucleotide or
CC polypeptide of the invention is useful for preventing, treating or
CC ameliorating a medical condition, e.g. a proliferative disorder. They are
CC also useful for treating e.g. liver disease, renal failure, immunologic
CC disorders including arthritis and psoriasis, cardiovascular disorders
CC such as congenital heart defects and congestive heart failure, and
CC cancer. A method of the invention is useful for diagnosing a pathological
CC condition or susceptibility to a condition in a subject. The present
CC sequence is used in the exemplification of the invention

SQ Sequence 665 AA;

Query Match	100.0%;	Score 665;	DB 5;	Length 665;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 665;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

QY	1	MAHEMIGTQVTERLVALLSEGTENKLLIDSPPEVNTSHILEAININCSKLMKRLLQQ	60
Dd	1	MAHEMIGTQVTERLVALLSEGTENKLLIDSPPEVNTSHILEAININCSKLMKRLLQQ	60
QY	61	DKVLITELLIOHSAKHVIDIDCGQRVVYVDQSODVASISPDCLFTVLGKEKS FNSVHL	120
Dd	61	DKVLITELLIOHSAKHVIDIDCGQRVVYVDQSODVASISPDCLFTVLGKEKS FNSVHL	120
QY	121	LAGPFAEFSCRCPGLCEGKSLTPVCISOPCLPVANIGPTBILPMLYLGGORDVLNLELI	180
Dd	121	LAGPFAEFSCRCPGLCEGKSLTPVCISOPCLPVANIGPTBILPMLYLGGORDVLNLELI	180
QY	181	OONGIGVYNASTYCKPDPFIESHFLRVPVDSFCFKILLPLDLSVDFIEKAASNGCV	240
Dd	181	OONGIGVYNASTYCKPDPFIESHFLRVPVDSFCFKILLPLDLSVDFIEKAASNGCV	240
QY	241	LVHCLAGISRSATIAIAYIMKMDMSLDAYR FVKEKX PTLSPNNPIGQLLDYEKKIKX	300
Dd	241	LVHCLAGISRSATIAIAYIMKMDMSLDAYR FVKEKX PTLSPNNPIGQLLDYEKKIKX	300
QY	301	QTGASGPYSKL KLIHLEKENEPVPAVSEGGQKSETPLSPCCADSATSEAAQORVHPASV	360
Dd	301	QTGASGPYSKL KLIHLEKENEPVPAVSEGGQKSETPLSPCCADSATSEAAQORVHPASV	360
QY	361	PSPVPSQPSSLBESPLVOALSGIHLSADRLEBSNLKRXSFSLIDIKSYSYASMAASHGF	420
Dd	361	PSPVPSQPSSLBESPLVOALSGIHLSADRLEBSNLKRXSFSLIDIKSYSYASMAASHGF	420
QY	421	SSEEDDALEYKPPSTTTDGTNKLCQFSPVOELSEOPTETSPDKEEASIPKLUQTARPSDSQ	480
Dd	421	SSEEDDALEYKPPSTTTDGTNKLCQFSPVOELSEOPTETSPDKEEASIPKLUQTARPSDSQ	480
QY	481	SKRLHSVRSSSGTAORSLISPLHRSGVEDNYHTSF LGLSTSQQHLYTKSAGLGKGWH	540
Dd	481	SKRLHSVRSSSGTAORSLISPLHRSGVEDNYHTSF LGLSTSQQHLYTKSAGLGKGWH	540
QY	541	SDILAPOTSPLTSGWFATESSHRYASA IYGGASAYSAYSCQLPTCGDOVYSVRRR	600
Dd	541	SDILAPOTSPLTSGWFATESSHRYASA IYGGASAYSAYSCQLPTCGDOVYSVRRR	600
QY	601	QKSDRADSRHSWHEESPFEKOFRKRCOMEGBEISMENSRREBLGVGSQSSFSGSME	660
Dd	601	QKSDRADSRHSWHEESPFEKOFRKRCOMEGBEISMENSRREBLGVGSQSSFSGSME	660
QY	661	IIEVS 665	
Dd	661	IIEVS 665	
<hr/>			
RESULT_3			
AAU79159			
ID	AAU79159 standard; protein; 517 AA.		
AC			
XX	AAU79159;		
DT	02-JUL-2002 (first entry)		

RESULT 3
AAU79159

XX Human dual-specificity phosphatase-3 (DSP-16) alternative form protein
XX
XX Human: dual-specificity phosphatase-3; DSP-16; autoimmune disease;
KW mitogen-activated protein kinase; MAP; Duchenne muscular dystrophy;
KW cancer; graft-versus-host disease; allergy; metabolic disease;
KW abnormal cell growth; abnormal cell proliferation; contract inhibition;
KW cell cycle abnormality; anchorage independent cell growth; apoptosis;
KW intercellular adhesion; DSP-16 modulator; chromosome 12p.

OS Homo sapiens.

	Key	Location/Qualifiers
FH	Domain	94. .103
FT		/label= Active_site_domain
FT		

PN WO200226997-A2

PD 04-APR-2002

PF 25-SEP-2001; 2001WO-US030124

PR 26-SEP-2000; 2000US-0235487P

PA (CEPT-) CEPTYR INC

PI Luche RM, Wei B,

DR WPI; 2002-315802/35

XX

PT New DSF-16 polypeptide, useful for identifying modulators of its
PT activity, which can be used in the treatment of disorders such as
PT Duchenne muscular dystrophy, or cancer.

PS Claim 50; Fig 4; 87pp; English

CC The present invention relates to a new polypeptide, DSP-16, having a 665
CC amino acid sequence, given in the specification, or a variant having at
CC least 50 % identical residues, which retains the ability to
CC dephosphorylate an activated mitogen-activated protein (MAP) kinase. The
CC invention can be used for identifying agents which modulate DSP-16
CC activity, for modulation of a proliferative response in a cell, survival
CC of a cell, or differentiation of a cell. The cell displays contact
CC inhibition of cell growth or anchorage independent growth and may display
CC altered intercellular adhesion. The agent may modulate apoptosis, or the
CC cell cycle. The identified modulators can be used to treat Duchenne
CC muscular dystrophy, cancer, graft-versus-host disease, autoimmune
CC diseases, allergies, metabolic diseases, abnormal cell growth, abnormal
CC cell proliferation, and cell cycle abnormalities. The present amino acid
CC sequence represents the human dual-specificity phosphatase-3 (DSP-16)
CC alternative form protein of the invention. This sequence is encoded by
CC the human DSP-16 gene located on chromosome 12p

Sequence 517 AA;

Query Match	71.0%	Score 472	DB 5	length 517
Best Local Similarity	100.0%	Pred. No. 0		
Matches 472	Conservative 0	Mismatches 0	Indels 0	Gaps 0

Oy	194	TCBPRDPLPESHFLVRVPNDSCFKILPYMLDKSDVTEKAKASNCVYHCLAGISRSAT	253
Db	46	TCBPRDPLPESHFLVRVPNDSCFKILPYMLDKSDVTEKAKASNCVYHCLAGISRSAT	105
Oy	254	IAIAYIMKRMMSLDEAYRFYKERRPTSPNPNFGOLLDYKIKIIONOTGASGPKSKYL	313
Db	106	IAIAYIMKRMMSLDEAYRFYKERRPTSPNPNFGOLLDYKIKIIONOTGASGPKSKYL	165
Oy	314	LHLEKPNRPVAVSEGGOKSETPLSPPCADSATSEAGQRVNHPRASVSVSVQPSLLED	373
Db	166	LHLEKPNRPVAVSEGGOKSETPLSPPCADSATSEAGQRVNHPRASVSVSVQPSLLED	225
Oy	374	SPVIOALSGLLSADRLEDNSKLRFSGLDIKSVSYASMAASLHGSSSBDALBYKPS	433

Db 226 SPLVQALSGHLASADRLSDSNKLRKSFSLDIKVSYSASMAASLHGFSSEDALEYKPS 285
QY 434 TTLDTGNKLCQFSPVOELSEQTPETSPDKEEASIPKQLQTARPSDSQSKRLHSVTTSSSG 493
Db 286 TTLDTGNKLCQFSPVOELSEQTPETSPDKEEASIPKQLQTARPSDSQSKRLHSVTTSSSG 345
QY 494 TQORSLSLPLHRSVGSVEDNYHTSFLFGLSTSQOHLTKSAGLGKGMHSDILAPOTSTPBL 553
Db 346 TQORSLSLPLHRSVGSVEDNYHTSFLFGLSTSQOHLTKSAGLGKGMHSDILAPOTSTPBL 405
QY 554 TSSWPFATSSHPYASAIYGGASAYSASCSQLPTCGOVYSVRRQKPSRADSRBM 613
Db 406 TSSWPFATSSHPYASAIYGGASAYSASCSQLPTCGOVYSVRRQKPSRADSRBM 465
QY 614 HESPEKQFKRRSCOMERGESIMSENRREELGKVGSGSFSFGSMETIEVS 665
Db 466 HESPEKQFKRRSCOMERGESIMSENRREELGKVGSGSFSFGSMETIEVS 517

RESULT 4
AAE04834
ID AAE04834 standard; protein; 665 AA.
AC AAE04834;
XX
XX
XX 10-SEP-2001 (first entry)
DE Human SGP002 phosphatase polypeptide.
XX
XX Human; SGP002 phosphatase polypeptide; phosphatase-related disease;
KW immune-related disorder; ocular disease; organ transplant rejection;
KW infection; diabetes; pain; sexual dysfunction; Alzheimer's disease;
KW metabolic disorder; haematopoietic cancer; mood disorder; cardiac;
KW Parkinson's disease; multiple sclerosis; amyotrophic lateral sclerosis;
KW cardiovascular disease; brain; neuronal-associated disease; dyskinesia;
KW attention disorder; cognition disorder; psychotic disorder; cytostatic;
KW neurological disorder; virulence; noctropic; cerebroprotective therapy;
KW neuroprotective; antibacterial; vulnerrary; tranquiliser; antischmatic;
KW hypotensive; immunosuppressive; antiporiatic; analgesic; hypertensive;
KW antifungal; dual specificity phosphatase; DSP; MAP kinase phosphatase;
KW MKP; migraine; chromosome 12p11.1-p12.1.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FT Domain 1..173
FT /label= Catalytic_domain
FT 158..297
FT Domain /label= Phosphatase_domain
XX
XX WO200146394-A2.
PN
XX
XX 28-JUN-2001.
PD
XX
XX 21-DEC-2000; 2000MO-US034736.
PF
XX
XX 21-DEC-1999; 99US-0173255P.
PR 28-DEC-1999; 99US-0175766P.
PR 25-JAN-2000; 2000US-0178078P.
PR 31-JAN-2000; 2000US-0179301P.
XX
XX (SUGEN) SUGEN INC.
XX
XX Plowman GD, Martinez R, Whyte D, Manning G, Sudarsanam S;
PI Hill RJ, Flanagan P;
XX
XX WPI: 2001-418058/44.
DR N-PSDB; AAD09492.
XX
XX Novel phosphatase polypeptide useful for treating cancers, immune-related
PT diseases and disorders; cardiovascular disease, brain or neuronal-
PT associated diseases and metabolic disorders.

XX
PS Claim 7; Fig 2; 186pp; English.
CC
CC The present invention relates to phosphatase polypeptides, nucleotide
CC sequences encoding them, as well as various products and methods useful
CC for the diagnosis and treatment of various phosphatase-related diseases
CC and conditions. Substance that modulates the activity of phosphatase
CC polypeptide is used to treat immune-related diseases and disorders,
CC cardiovascular disease, brain or neuronal-associated diseases and
CC metabolic disorders, including cancers of tissues, cancers of
CC haematopoietic origin, diseases of central and peripheral nervous system,
CC Alzheimer's disease, Parkinson's disease, multiple sclerosis, amyotrophic
CC lateral sclerosis, viral infections, infections caused by prions,
CC bacteria and fungi, ocular diseases, diabetes, migraines, pain, sexual
CC dysfunction, mood disorders, attention disorders, cognition disorders,
CC hypertension, hyperextension, psychotic disorders, neurological disorders,
CC dyskinesias and organ transplant rejection. The present amino acid
CC sequence is human SGP002 phosphatase polypeptide. This sequence is
CC classified as dual specificity phosphatase (DSP) and MAP kinase
CC phosphatase (MKP). SGP002 gene maps to chromosomal position 12p11.1-p12.1
XX
SQ Sequence 665 AA;

Query Match 71.0%; Score 472; DB 4; Length 665;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 TCPKPDPIPESHFLAVPVNDSECEKILPWLDKSVDFIKAKASNCVUHLAIGSRAT 253
Db 194 TCPKPDPIPESHFLAVPVNDSECEKILPWLDKSVDFIKAKASNCVUHLAIGSRAT 253
QY 254 IAIAYIMKMDMSLDEAYRFEVEKRPISPNFPGQLLDYEKKIKNOTGASGPKSLKT 313
Db 254 IAIAYIMKMDMSLDEAYRFEVEKRPISPNFPGQLLDYEKKIKNOTGASGPKSLKT 313
QY 314 IHLKRNPEVPAVSEGGKSETPLSPCADSATSAACQRPVHPASVPSVPSVQSLLED 373
Db 314 IHLKRNPEVPAVSEGGKSETPLSPCADSATSAACQRPVHPASVPSVPSVQSLLED 373
QY 374 SPLVQALSGHLASADRLSDSNKLRKSFSLDIKVSYSASMAASLHGFSSEDALEYKPS 433
Db 374 SPLVQALSGHLASADRLSDSNKLRKSFSLDIKVSYSASMAASLHGFSSEDALEYKPS 433
QY 434 TTLDTGNKLCQFSPVOELSEQTPETSPDKEEASIPKQLQTARPSDSQSKRLHSVTTSSSG 493
Db 434 TTLDTGNKLCQFSPVOELSEQTPETSPDKEEASIPKQLQTARPSDSQSKRLHSVTTSSSG 493
QY 494 TQORSLSLPLHRSVGSVEDNYHTSFLFGLSTSQOHLTKSAGLGKGMHSDILAPOTSTPBL 553
Db 494 TQORSLSLPLHRSVGSVEDNYHTSFLFGLSTSQOHLTKSAGLGKGMHSDILAPOTSTPBL 553
QY 554 TSSWPFATSSHPYASAIYGGASAYSASCSQLPTCGOVYSVRRQKPSRADSRBM 613
Db 554 TSSWPFATSSHPYASAIYGGASAYSASCSQLPTCGOVYSVRRQKPSRADSRBM 613
QY 614 HESPEKQFKRRSCOMERGESIMSENRREELGKVGSGSFSFGSMETIEVS 665
Db 614 HESPEKQFKRRSCOMERGESIMSENRREELGKVGSGSFSFGSMETIEVS 665

RESULT 5
AAU09016
ID AAU09016 standard; protein; 665 AA.
AC AAU09016;
XX
XX 18-DEC-2001 (first entry)
DT
XX
XX Human dual specificity phosphatase 21117.
DE
XX
XX Human; dual specificity phosphatase 21117; hepatotropic; cytostatic;
KW hematopoietic disorder; autoimmune disorder; diabetes mellitus;
KW rheumatoid arthritis; multiple sclerosis; Crohn's disease;

KW liver disorder; erythroid associated disorder; haemolytic anaemia;
 KW cellular proliferative; differentiative disorder; leukaemia;
 KW acute myeloid leukaemia; carcinoma; sarcoma; metastatic cancer;
 KW immunogen.
 XX Homo sapiens.
 OS
 XX
 XX Key Location/Qualifiers
 PH Domain 11..131
 FT Domain /label= Rhodanese-like domain
 FT 158..297
 FT /label= Catalytic domain
 FT /note= "Dual specificity phosphatase catalytic domain"
 FT 242..254
 FT Active-site /label= Tyrosine-specific_protein_phosphatase_active_site
 FT
 PN WO200173059-A2.
 XX
 XX 04-OCT-2001.
 PD
 XX
 XX 23-MAR-2001; 2001WO-US009477.
 PF
 XX
 XX 24-MAR-2000; 2000US-0191858P.
 PR
 XX
 XX (MILL-) MILLENNIUM PHARM INC.
 PA
 XX
 PI Meyers RA;
 XX
 DR WPI; 2001-611635/70.
 DR N-PSDB; AAS14639.
 XX
 XX
 PT New human dual specificity polypeptides and nucleic acids for diagnosis
 of disease and treatment of e.g. liver disorders.
 XX
 XX
 XX Claim 9; Fig 1; 143pp; English.
 XX
 XX The invention relates to two novel human dual specificity phosphatases
 designated 2117 and 38692, the nucleic acids encoding them (including
 fragments, allelic variants, their complements or nucleic acids that
 hybridize to them) and antibodies raised against the proteins. The
 CC antibody is useful for detecting the presence of the polypeptide, and the
 CC nucleic acid fragments are useful for detecting the presence of the
 CC nucleic acid, for diagnosing disease. Agents (e.g. the antibody and
 CC antisense sequences) are useful for modulating the activity or expression
 of the polypeptide/nucleic acid, useful for the treatment of e.g.
 CC hematopoietic disorders, autoimmune disorders (e.g. diabetes mellitus,
 CC rheumatoid arthritis, multiple sclerosis, Crohn's disease and many others
 CC listed in the specification) liver disorders, erythroid associated
 CC disorders (e.g. haemolytic anaemia) cellular proliferative or
 CC differentiative disorders, leukaemias (e.g. acute myeloid leukaemia),
 CC carcinomas, sarcomas, metastatic cancers. Agents modulating 2117 and
 CC 38692 are also useful for modulating the proliferation, survival,
 CC migration or differentiation of a 38692 or 2117-expressing cell. The
 CC polypeptide and nucleic acids are useful for identifying modulating
 CC agents. The present sequence represents the dual specificity phosphatase
 CC 2117
 XX
 XX
 XX Sequence 665 AA;
 SQ
 Query Match 71.0%; Score 472; DB 4; Length 665;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 194 TCCKPDPFIPESHFLRVAVNDSPCEKILPMWDKSVDFIEKAKSNGCVLVHCLAGISRAT 253
 DB 194 TCCKPDPFIPESHFLRVAVNDSPCEKILPMWDKSVDFIEKAKSNGCVLVHCLAGISRAT 253
 QY 254 IAAIYIMKRMDSIDEXYRVKKEKRPISNPNFLGQLDYEKIKKQGTASPKSKLKL 313
 DB 254 IAAIYIMKRMDSIDEXYRVKKEKRPISNPNFLGQLDYEKIKKQGTASPKSKLKL 313
 QY 314 LHLKXNEPVPVASEGQKSETPLSPCADSATSEAGORPVVPASVSPVQPSILSD 373

DB 314 LHLKXNEPVPVASEGQKSETPLSPCADSATSEAGORPVVPASVSPVQPSILSD 373
 QY 374 SPLVQALSGHLHSARLSDSNLKRFSFLDIKSVSYASMAASLIGFSSSEDALEYKPS 433
 DB 374 SPLVQALSGHLHSARLSDSNLKRFSFLDIKSVSYASMAASLIGFSSSEDALEYKPS 433
 QY 434 TTLDTGNKLCQFSPVQELSEQTPETSPDKEASIRPKLQTAAPSDSQSKRLSVRTSSG 493
 DB 434 TTLDTGNKLCQFSPVQELSEQTPETSPDKEASIRPKLQTAAPSDSQSKRLSVRTSSG 493
 QY 494 TQORSILSPILHRSGSVEDNYHTSFLGLSTSQQLTKSAGLCKGMHSDILAPQSTPBL 553
 DB 494 TQORSILSPILHRSGSVEDNYHTSFLGLSTSQQLTKSAGLCKGMHSDILAPQSTPBL 553
 QY 554 TSSWYFATSSHFFVSASAIYGGSSASVSCGLPTCCGQVSVRRQRPSPRADSRSG 613
 DB 554 TSSWYFATSSHFFVSASAIYGGSSASVSCGLPTCCGQVSVRRQRPSPRADSRSG 613
 QY 614 HESPEKQFKRSCOMERFESIMSENRREELGKVSQSPSGMEITEVS 665
 DB 614 HESPEKQFKRSCOMERFESIMSENRREELGKVSQSPSGMEITEVS 665
 RESULT 6
 AAU79156
 ID AAU79156 standard; protein; 665 AA.
 AC AAU79156;
 XX
 XX
 DT 02-JUL-2002 (first entry)
 XX
 XX Human dual-specificity phosphatase-3 (DSP-16) protein.
 DE
 XX
 XX Human; dual-specificity phosphatase-3; DSP-16; autoimmune disease;
 KW mitogen-activated protein kinase; MAP; Duchenne muscular dystrophy;
 KW cancer; graft-versus-host disease; allergy; metabolic disease;
 KW abnormal cell growth; abnormal cell proliferation; contact inhibition;
 KW cell cycle abnormality; anchorage independent cell growth; apoptosis;
 KW intercellular adhesion; DSP-16 modulator; chromosome 12p.
 XX
 XX
 OS Homo sapiens.
 XX
 XX
 FH Key Location/Qualifiers
 FT Domain 242..251
 FT /label= Active_site_domain
 PN WO200226997-A2.
 XX
 PD 04-APR-2002.
 XX
 XX 25-SEP-2001; 2001WO-US030124.
 PF
 XX
 PR 26-SEP-2000; 2000US-0235487P.
 XX
 XX
 PA (CEPT-) CEPTYR INC.
 XX
 XX
 PI Lucine RM, Wei B;
 XX
 DR WPI; 2002-315802/35.
 DR N-PSDB; ABK47596.
 XX
 XX
 PT New DSP-16 polypeptide, useful for identifying modulators of its
 PT activity, which can be used in the treatment of disorders such as
 PT Duchenne muscular dystrophy, or cancer.
 PS
 XX
 XX Claim 1; Fig 2; 87pp; English.
 CC The present invention relates to a new polypeptide, DSP-16, having a 665
 CC amino acid sequence, given in the specification, or a variant having at
 CC least 50 % identical residues, which retains the ability to
 CC phosphorylate an activated mitogen-activated protein (MAP) kinase. The
 CC invention can be used for identifying agents which modulate DSP-16
 CC activity, for modulation of a proliferative response in a cell, survival

CC of a cell, or differentiation of a cell. The cell displays contact
CC inhibition of cell growth or anchorage independent growth and may display
CC altered intercellular adhesion. The agent may modulate apoptosis, or the
CC cell cycle. The identified modulators can be used to treat Duchenne
CC muscular dystrophy, cancer, graft-versus-host disease, autoimmune
CC diseases, allergies, metabolic diseases, abnormal cell growth, abnormal
CC cell proliferation, and cell cycle abnormalities. The present amino acid
CC sequence represents the human dual-specificity phosphatase-3 (DSP-16)
CC protein of the invention. This sequence is encoded by the human DSP-16
CC gene located on chromosome 12p
XX
XX
SQ Sequence 665 AA;

Query Match 71.0%; Score 472; DB 5; Length 665;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 TCPKPDPIESHFLRVVNDSPCEKILPMIDKSVDFIEKAKASNGCVLVHCLAGISRAT 253
DB 194 TCPKPDPIESHFLRVVNDSPCEKILPMIDKSVDFIEKAKASNGCVLVHCLAGISRAT 253
QY 254 IAIAYIMKMDMSLDEAYRFVKEKRPPTISPNFNFLGQLLDYEKKIKNOTGASGPKSKL 313
DB 254 IAIAYIMKMDMSLDEAYRFVKEKRPPTISPNFNFLGQLLDYEKKIKNOTGASGPKSKL 313
QY 314 IHLKPNRPVPAVSEGGOKSETPPLSPCADSATSEAAAGRPVHPASVPSVPSQLLED 373
DB 314 IHLKPNRPVPAVSEGGOKSETPPLSPCADSATSEAAAGRPVHPASVPSVPSQLLED 373
QY 374 SPLVQALSGHLASADRLSDSNLKRFSFLDIKSVSYASMAASLHGFSSSEDALEYKPS 433
DB 374 SPLVQALSGHLASADRLSDSNLKRFSFLDIKSVSYASMAASLHGFSSSEDALEYKPS 433
QY 434 TTLDTGNKLCQPSVQELSEQTPETSPDXEASIPKKLOTARPDSQSKRLHSVTSSSG 493
DB 434 TTLDTGNKLCQPSVQELSEQTPETSPDXEASIPKKLOTARPDSQSKRLHSVTSSSG 493
QY 494 TAQRSLSLPLHRSGVEDVNYHTSFLGLSTSOOHLTKSAGLGLKGMHSDIAPOTSTPSL 553
DB 494 TAQRSLSLPLHRSGVEDVNYHTSFLGLSTSOOHLTKSAGLGLKGMHSDIAPOTSTPSL 553
QY 554 TSSWYFATESSHFYASAIYGGASAYSAYSCQLPTCGDQVYSVRRROKPSDRADSRSM 613
DB 554 TSSWYFATESSHFYASAIYGGASAYSAYSCQLPTCGDQVYSVRRROKPSDRADSRSM 613
QY 614 HESPEKQFKRRSCQMEFGESIMSENRSREELGKVGSSQSFSGSMEIIEVS 665
DB 614 HESPEKQFKRRSCQMEFGESIMSENRSREELGKVGSSQSFSGSMEIIEVS 665

RESULT 7
AAU09946 standard; protein; 665 AA.
ID AAU09946;
AC AAU09946;
XX
XX
DT 18-JUN-2002 (first entry)
XX
XX
DE Protein sequence of human (dual specificity phosphatase) DSP-10.
XX
XX
KW Human; dual specificity phosphatase; DSP-10; cancer; epilepsy; stroke;
KW neuronal degeneration syndrome; Alzheimer's disease; depression;
KW schizophrenia; asthma; immune disorder; inflammatory process; arthritis;
KW osteoporosis; diabetes.
XX
XX
OS Homo sapiens.
XX
PN MO200177340-A1.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-EP003966.
XX

PR 10-APR-2000; 2000EP-00107143.
XX
XX (MERE) MERCK PATENT GMEH.
PA
XX
XX
PI Duecker K;
XX
XX
DR MPI: 2002-010917/01.
DR N-PsDB; MA515768.
XX
XX
PT Novel dual specificity phosphatase polypeptides useful for treating
PT cancers, epilepsy, stroke, neuronal degeneration syndromes, Alzheimer's
PT disease, depression, schizophrenia, asthma and immune disorders.
XX
XX
PS Claim 2; Page 37-39; 43pp; English.

CC The present invention relates to a new isolated dual specificity
CC phosphatase (DSP10) polypeptide, comprising a 665 residue amino acid
CC sequence that is fully defined in the specification. The invention also
CC provides a sequence encoded by a 3059 nucleotide sequence fully defined
CC in the specification, and a sequence having at least 95 % identity to the
CC polypeptide, or fragments or variants of DSP-10. The invention is useful
CC for treating cancer e.g. leukemia, colon carcinoma, lung cancer,
CC prostate cancer, metastasis of tumour cells, neo-angiogenesis, epilepsy,
CC stroke, neuronal degeneration syndromes, Alzheimer's disease, depression,
CC schizophrenia, cardiac myopathies, asthma, immune disorders,
CC inflammatory processes e.g. arthritis, bowel disease, type I diabetes,
CC osteoporosis, diabetes and diabetes associated diseases. The molecules of
CC the invention are also useful as vaccines for inducing immunological
CC response in a mammal, in disease diagnosis and in assays for screening
CC agonistic or antagonistic compounds. Other uses of the invention include
CC identifying membrane bound or soluble receptors, as a diagnostic reagent,
CC in chromosome localisation studies, and as a valuable tool in tissue
CC expression studies. The present sequence represents the dual specificity
CC phosphatase, DSP-10, protein of the invention
XX
XX
SQ Sequence 665 AA;

Query Match 71.0%; Score 472; DB 5; Length 665;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 TCPKPDPIESHFLRVVNDSPCEKILPMIDKSVDFIEKAKASNGCVLVHCLAGISRAT 253
DB 194 TCPKPDPIESHFLRVVNDSPCEKILPMIDKSVDFIEKAKASNGCVLVHCLAGISRAT 253
QY 254 IAIAYIMKMDMSLDEAYRFVKEKRPPTISPNFNFLGQLLDYEKKIKNOTGASGPKSKL 313
DB 254 IAIAYIMKMDMSLDEAYRFVKEKRPPTISPNFNFLGQLLDYEKKIKNOTGASGPKSKL 313
QY 314 IHLKPNRPVPAVSEGGOKSETPPLSPCADSATSEAAAGRPVHPASVPSVPSQLLED 373
DB 314 IHLKPNRPVPAVSEGGOKSETPPLSPCADSATSEAAAGRPVHPASVPSVPSQLLED 373
QY 374 SPLVQALSGHLASADRLSDSNLKRFSFLDIKSVSYASMAASLHGFSSSEDALEYKPS 433
DB 374 SPLVQALSGHLASADRLSDSNLKRFSFLDIKSVSYASMAASLHGFSSSEDALEYKPS 433
QY 434 TTLDTGNKLCQPSVQELSEQTPETSPDXEASIPKKLOTARPDSQSKRLHSVTSSSG 493
DB 434 TTLDTGNKLCQPSVQELSEQTPETSPDXEASIPKKLOTARPDSQSKRLHSVTSSSG 493
QY 494 TAQRSLSLPLHRSGVEDVNYHTSFLGLSTSOOHLTKSAGLGLKGMHSDIAPOTSTPSL 553
DB 494 TAQRSLSLPLHRSGVEDVNYHTSFLGLSTSOOHLTKSAGLGLKGMHSDIAPOTSTPSL 553
QY 554 TSSWYFATESSHFYASAIYGGASAYSAYSCQLPTCGDQVYSVRRROKPSDRADSRSM 613
DB 554 TSSWYFATESSHFYASAIYGGASAYSAYSCQLPTCGDQVYSVRRROKPSDRADSRSM 613
QY 614 HESPEKQFKRRSCQMEFGESIMSENRSREELGKVGSSQSFSGSMEIIEVS 665
DB 614 HESPEKQFKRRSCQMEFGESIMSENRSREELGKVGSSQSFSGSMEIIEVS 665

```

RESULT 8
AAU75789
ID AAU75789 standard; protein; 665 AA.
AC
XX AAU75789;
XX
DT 08-MAY-2002 (first entry)
XX
XX Human protein phosphatase 7 (PP7) protein sequence.
XX
XX Human; protein phosphatase 7; PP7; immune system disorder; AIDS;
XX acquired immunodeficiency syndrome; thymic hypoplasia; anaemia; asthma;
XX Crohn's disease; neurological disorder; epilepsy; Huntington's disease;
XX dementia; Parkinson's disease; developmental disorder; Down's syndrome;
XX cell proliferative disorder; cancer; adenocarcinoma; leukaemia; lymphoma;
XX melanoma; myeloma sarcoma.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX Domain 11..131
XX /label= Rhodanese_like_domain
XX FT 15..170
XX /note="Dual specificity protein phosphatase"
XX FT 85..298
XX /note="WH1-type dual specificity protein phosphatase"
XX FT 158..297
XX /label= Catalytic domain DSPc
XX FT /note="Dual specificity phosphatase"
XX FT 220..281
XX /note="Tyrosine specific protein phosphatases active
XX FT site tyr_phosphatase.ptf"
XX FT 242..254
XX /note="Tyrosine phosphatase"
XX FT 242..252
XX /note="Tyrosine specific protein phosphatases"
XX PN MO200210363-A2.
XX
XX 07-FEB-2002.
XX
XX 26-JUL-2001; 2001WO-US023716.
XX
XX 28-JUL-2000; 2000US-0221679P.
XX PR 03-AUG-2000; 2000US-0223272P.
XX PR 10-AUG-2000; 2000US-0224309P.
XX PR 18-AUG-2000; 2000US-0226728P.
XX PR 30-AUG-2000; 2000US-0229254P.
XX PR 08-SEP-2000; 2000US-0231366P.
XX
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Tang YT, Elliott VS, Ramkumar J, Yao MG, Burford N, Wang YE;
XX Stewart EA, Gandhi AR, Patterson C, Lee EA, Hafalia ADA, Lu DM;
XX Tribouley CM, Griffin JA, Baughn MR, Yue H, Warren BA, Nguyen DB;
XX Walla NK, Kearney L;
XX
XX WPI; 2002-188735/24.
XX DR N-PSDB; ABK14474.
XX
XX New protein phosphatases, useful for diagnosing, treating or preventing
XX immune system disorders (e.g. Crohn's disease), neurological disorders
XX (e.g. Parkinson's disease), or cell proliferative disorders (e.g.
XX cancers).
XX
XX Claim 1; Page 107-108; 117pp; English.
XX
XX The present invention relates to a new polypeptide, a naturally occurring
XX amino acid sequence at least 95 % identical to it, a biologically active
XX fragment of it or an immunogenic fragment of it. The polypeptides,
XX polynucleotides, agonists and antagonists are useful for diagnosing,
XX treating or preventing disorders associated with aberrant expression of

```

```

CC protein phosphatases (PP), particularly immune system disorders e.g.
CC acquired immunodeficiency syndrome (AIDS), thymic hypoplasia, anaemia,
CC asthma or Crohn's disease, neurological disorders e.g. epilepsy,
CC Huntington's disease, dementia or Parkinson's disease, developmental
CC disorders e.g. Down's syndrome, or cell proliferative disorders e.g.
CC cancers including adenocarcinoma, leukaemia, lymphoma, melanoma, myeloma
CC or sarcoma. The present amino acid sequence represents human protein
CC phosphatase 7 (PP7) which is one of several human protein phosphatases
CC (AAU75783-AAU75792) of the invention
XX
XX Sequence 665 AA;
XX
XX Query Match 71.0%; Score 472; DB 5; Length 665;
XX Best Local Similarity 100.0%; Pred NC 0;
XX Matches 4/72; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 194 TCEPKDFIPESHFLRPVNDSECEKILPWLKDSVDEIEKAKANGCVLHCLAGISRSAT 253
XX DB 194 TCEPKDFIPESHFLRPVNDSECEKILPWLKDSVDEIEKAKANGCVLHCLAGISRSAT 253
XX
XX QY 254 IAIAYIMKMDMSLDEAYRFVYKERTTISPNFNLGQLLDYEKKIKNOTGASGPKSKLXL 313
XX DB 254 IAIAYIMKMDMSLDEAYRFVYKERTTISPNFNLGQLLDYEKKIKNOTGASGPKSKLXL 313
XX
XX QY 314 LHEKNEBPVAVSBGGQKSETPLSPCADSATSEAAAGRPVHPASVPSPVQPSLLEP 373
XX DB 314 LHEKNEBPVAVSBGGQKSETPLSPCADSATSEAAAGRPVHPASVPSPVQPSLLEP 373
XX
XX QY 374 SPLVQALSGHLMSADRLSDSNLKLKSFSLDIKSVSYASMAAHLHGSSSEDLLEYKPS 433
XX DB 374 SPLVQALSGHLMSADRLSDSNLKLKSFSLDIKSVSYASMAAHLHGSSSEDLLEYKPS 433
XX
XX QY 434 TTLDTGNKLCOFSPVOELSEQTPETSPPDEKASIPKKLOTAPPSDQSRLHSVRTSSG 493
XX DB 434 TTLDTGNKLCOFSPVOELSEQTPETSPPDEKASIPKKLOTAPPSDQSRLHSVRTSSG 493
XX
XX QY 494 TQRSLSLPLHRSQSVEDNYHTSFLGLSTSOQHLTKSAGLKGWHSIDLAPQSTPPL 553
XX DB 494 TQRSLSLPLHRSQSVEDNYHTSFLGLSTSOQHLTKSAGLKGWHSIDLAPQSTPPL 553
XX
XX QY 554 TSSWYFATESSHFYSAIYGSASAYSAYSCQLPTCGDQVYSVRRRQKPSDRADRRSW 613
XX DB 554 TSSWYFATESSHFYSAIYGSASAYSAYSCQLPTCGDQVYSVRRRQKPSDRADRRSW 613
XX
XX QY 614 HESPEPKQFKRRSCMEFGEISIMENRREELGKYGSSGSPSGMEITEVS 665
XX DB 614 HESPEPKQFKRRSCMEFGEISIMENRREELGKYGSSGSPSGMEITEVS 665
XX
XX RESULT 9
XX ABB97946
XX ID ABB97946 standard; protein; 665 AA.
XX
XX AC ABB97946;
XX
XX DT 06-SEP-2002 (first entry)
XX
XX Human protein sequence #13.
XX
XX Human; brain; tonsil; hippocampus; foetal brain; diagnosis.
XX
XX Homo sapiens.
XX
XX WO200252005-A1.
XX
XX 04-JUL-2002.
XX
XX 20-DEC-2001; 2001WO-JP011217.
XX
XX 22-DEC-2000; 2000JP-00389742.
XX
XX (KAZU-) KAZUSA DNA RBS INST FOUND.
XX
XX (CELE-) CELESTAR LEXICO-SCI LTD.

```


XX SQ Sequence 665 AA;
Query Match 71.0%; Score 472; DB 5; Length 665;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 194 TCPKPFIPESHFLRPVNDSPCEKILPWLDKSVDFIEKAKANGCVLHCLAGISRSAT 253
DB 194 TCPKPFIPESHFLRPVNDSPCEKILPWLDKSVDFIEKAKANGCVLHCLAGISRSAT 253
OY 254 IAIAYIMKMDMSLDEAYRFVKEKPTTISPNNFLGQLDYEKKIKNQTGASGPKSKLKL 313
DB 254 IAIAYIMKMDMSLDEAYRFVKEKPTTISPNNFLGQLDYEKKIKNQTGASGPKSKLKL 313
OY 314 LHLEKNEBPVAVSEGGOKSETPLSPPCADSATSEAGORPVHPASVPSVQPSLLED 373
DB 314 LHLEKNEBPVAVSEGGOKSETPLSPPCADSATSEAGORPVHPASVPSVQPSLLED 373
OY 374 SPLVQALSGHLASADRLSDSNLKRSPSLDIKSVSYASMAASLHGFSSEDLLEYKPS 433
DB 374 SPLVQALSGHLASADRLSDSNLKRSPSLDIKSVSYASMAASLHGFSSEDLLEYKPS 433
OY 434 TTLDTGNKLCQSPVOELSEQTPETSPDKEASIPKKLQTARPSDSQSKRLHSVRTSSG 493
DB 434 TTLDTGNKLCQSPVOELSEQTPETSPDKEASIPKKLQTARPSDSQSKRLHSVRTSSG 493
OY 494 TAORSLSLPLHRSQSVEDNYHTSFLGLSTSQOHLTKSAGLGKGMHSDILAPQISTPSL 553
DB 494 TAORSLSLPLHRSQSVEDNYHTSFLGLSTSQOHLTKSAGLGKGMHSDILAPQISTPSL 553
OY 554 TSSWYFATESSHFYASAIYGSASYSAYSCQLPTCGQVYSVRRQKPSDRADSRSM 613
DB 554 TSSWYFATESSHFYASAIYGSASYSAYSCQLPTCGQVYSVRRQKPSDRADSRSM 613
OY 614 HESPEKQFKRRSCQMEFGESIMSENRREBELGKVGSGSFGSGMEIIEVS 665
DB 614 HESPEKQFKRRSCQMEFGESIMSENRREBELGKVGSGSFGSGMEIIEVS 665
RESULT 11
ID ABB97291 standard; protein; 665 AA.
XX ABB97291;
XX 27-JUN-2002 (first entry)
XX DE Novel human protein SEQ ID NO: 559.
XX Human; antihaemic; vulnerary; antiinflammatory; immunomodulator;
XX antileukemia; cerebroprotective; cytoskeletal; rheumatic; gene therapy;
XX neuroprotective; antiParkinsonian; protein therapy; EST;
XX expressed sequence tag.
XX Homo sapiens.
XX WO20022660-A2.
XX 21-MAR-2002.
XX PF 10-SEP-2001; 2001WO-US026015.
XX PR 11-SEP-2000; 2000US-00659671.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
XX Xue AJ, Yang Y, Weinman T, Drmanac RT;
XX WPI; 2002-292408/33.
XX DR N-PSDB; ABN32477.
XX

PT An isolated polynucleotide for treating diseases associated with its
PT encoded polypeptide such as cancer and multiple sclerosis.
XX Example 2; SEQ ID NO 559; 509bp; English.
XX The present invention provides the protein and coding sequences of 444
XX novel human proteins. These were isolated from expressed sequences tags
XX (ESTs). They can be used to stimulate cell growth, to regulate
XX haematopoiesis e.g. to treat aplastic anemia, to help tissue regrowth
XX e.g. in burn treatment, to regulate the immune system e.g. to treat
XX multiple sclerosis, to regulate activin or inhibin e.g. to treat
XX infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke
XX and cancer, to screen for drugs, to treat inflammatory conditions e.g.
XX rheumatoid arthritis, and to treat nervous system disorders e.g.
XX Parkinson's disease. The present sequence is a protein of the invention
XX SQ Sequence 665 AA;
Query Match 71.0%; Score 472; DB 5; Length 665;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 194 TCPKPFIPESHFLRPVNDSPCEKILPWLDKSVDFIEKAKANGCVLHCLAGISRSAT 253
DB 194 TCPKPFIPESHFLRPVNDSPCEKILPWLDKSVDFIEKAKANGCVLHCLAGISRSAT 253
OY 254 IAIAYIMKMDMSLDEAYRFVKEKPTTISPNNFLGQLDYEKKIKNQTGASGPKSKLKL 313
DB 254 IAIAYIMKMDMSLDEAYRFVKEKPTTISPNNFLGQLDYEKKIKNQTGASGPKSKLKL 313
OY 314 LHLEKNEBPVAVSEGGOKSETPLSPPCADSATSEAGORPVHPASVPSVQPSLLED 373
DB 314 LHLEKNEBPVAVSEGGOKSETPLSPPCADSATSEAGORPVHPASVPSVQPSLLED 373
OY 374 SPLVQALSGHLASADRLSDSNLKRSPSLDIKSVSYASMAASLHGFSSEDLLEYKPS 433
DB 374 SPLVQALSGHLASADRLSDSNLKRSPSLDIKSVSYASMAASLHGFSSEDLLEYKPS 433
OY 434 TTLDTGNKLCQSPVOELSEQTPETSPDKEASIPKKLQTARPSDSQSKRLHSVRTSSG 493
DB 434 TTLDTGNKLCQSPVOELSEQTPETSPDKEASIPKKLQTARPSDSQSKRLHSVRTSSG 493
OY 494 TAORSLSLPLHRSQSVEDNYHTSFLGLSTSQOHLTKSAGLGKGMHSDILAPQISTPSL 553
DB 494 TAORSLSLPLHRSQSVEDNYHTSFLGLSTSQOHLTKSAGLGKGMHSDILAPQISTPSL 553
OY 554 TSSWYFATESSHFYASAIYGSASYSAYSCQLPTCGQVYSVRRQKPSDRADSRSM 613
DB 554 TSSWYFATESSHFYASAIYGSASYSAYSCQLPTCGQVYSVRRQKPSDRADSRSM 613
OY 614 HESPEKQFKRRSCQMEFGESIMSENRREBELGKVGSGSFGSGMEIIEVS 665
DB 614 HESPEKQFKRRSCQMEFGESIMSENRREBELGKVGSGSFGSGMEIIEVS 665
RESULT 12
ID ADA54744 standard; protein; 665 AA.
XX ADA54744;
XX 20-NOV-2003 (first entry)
XX DE Human protein, SEQ ID 2312.
XX Cytostatic; Anti-inflammatory; Osteopathic; Neuroprotective; Nocotropic;
XX Gene Therapy; human; secretory protein; membrane protein; cancer;
XX inflammatory disease; osteoporosis; neurological disease.
XX Homo sapiens.
XX EPI293569-A2.
XX

PD 19-MAR-2003.
 XX
 XX 21-MAR-2002; 2002EP-00006586.
 XX
 XX 14-SEP-2001; 2001JP-00328381.
 PR 24-JAN-2002; 2002US-0350435P.
 XX
 XX (HELI-) HELIX RES INST.
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.
 XX
 XX Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,
 XI Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuho Y;
 XX
 DR WPI; 2003-395539/38.
 DR N-PSDB; ADA53105.
 XX
 PT New polynucleotides encoding full-length polypeptides, e.g. secretory
 PT and/or membrane proteins, useful for developing medicines for diseases in
 PT which the gene is involved, or as target molecules for gene therapy.
 XX
 PS Claim 14; SEQ ID NO 2312; 205pp; English.
 XX
 CC The present invention relates to novel human secretory or membrane
 CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
 CC ADA54071). The coding sequences are useful in the gene therapy of
 CC diseases caused by abnormalities of the proteins, e.g. cancer,
 CC inflammatory diseases, osteoporosis or neurological disease.
 CC
 XX Sequence 665 AA;
 SQ

Query Match 71.0%; Score 472; DB 6; Length 665;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 TCEKPPFIESHFLRPVNDSPCEKILPMLDKSVDEIEKAKSNGCVLVHCLAGISRSAT 253
 DB 194 TCEKPPFIESHFLRPVNDSPCEKILPMLDKSVDEIEKAKSNGCVLVHCLAGISRSAT 253
 QY 254 IAIAYIMKMDMLDEAYRFVYKRPRTISNPNFLOQLDYEKKIKNOGASGPKSKKL 313
 DB 254 IAIAYIMKMDMLDEAYRFVYKRPRTISNPNFLOQLDYEKKIKNOGASGPKSKKL 313
 QY 314 IHLKENEVPVAVSEGGQKSETPLSPCADSATSEAGQRPVPAVSPVSPVQPSLLSD 373
 DB 314 IHLKENEVPVAVSEGGQKSETPLSPCADSATSEAGQRPVPAVSPVSPVQPSLLSD 373
 QY 374 SPLVQALSGIHLTADRLSDSNKLRKSFSLDIKSVSYASMAASLHGFSSEDALEYKPS 433
 DB 374 SPLVQALSGIHLTADRLSDSNKLRKSFSLDIKSVSYASMAASLHGFSSEDALEYKPS 433
 QY 434 TTLDTGNKLCQSPVQVDELSTQETSPDKKEASIPKKLQTPARSDSQSKLHVRTSSSG 493
 DB 434 TTLDTGNKLCQSPVQVDELSTQETSPDKKEASIPKKLQTPARSDSQSKLHVRTSSSG 493
 QY 494 TAQRSLSPILHRSGSVEDNHTSFLFGLSTQOHLTKSAGLKGMSHDLIAPQSTPSL 553
 DB 494 TAQRSLSPILHRSGSVEDNHTSFLFGLSTQOHLTKSAGLKGMSHDLIAPQSTPSL 553
 QY 554 TSSWYFATESSHFYSASAIYGSASYSAYSCSLPTCGDQVYVRRRKQKPSDRADRSRW 613
 DB 554 TSSWYFATESSHFYSASAIYGSASYSAYSCSLPTCGDQVYVRRRKQKPSDRADRSRW 613
 QY 614 HESSPFKQKRRSCQMEGESITMSNRREELGKVSQSSPFGSMIIIVS 665
 DB 614 HESSPFKQKRRSCQMEGESITMSNRREELGKVSQSSPFGSMIIIVS 665
 RESULT 13
 ID AAU79161 standard; protein: 665 AA.
 XX
 AC AAU79161;

XX 02-JUN-2002 (first entry)
 DT
 DE Human dual-specificity phosphatase-3 (DSP-16) mutant protein #1.
 XX
 XX Human; dual-specificity phosphatase-3; DSP-16; autoimmune disease;
 KW mitogen-activated protein kinase; MAP; Duchenne muscular dystrophy;
 KW cancer; graft-versus-host disease; allergy; metabolic disease;
 KW abnormal cell growth; abnormal cell proliferation; contact inhibition;
 KW cell cycle abnormality; anchorage independent cell growth; apoptosis;
 KW intercellular adhesion; DSP-16 modulator; mutant; mutein.
 XX
 OS Homo sapiens.
 OS Synthetic.
 OS
 FH Key location/Qualifiers
 FT Misc-difference 213
 FT /note= "Wild-type Asp substituted by Ala"
 XX
 PN MO200226997-AZ.
 XX
 PD 04-APR-2002.
 XX
 PF 25-SEP-2001; 2001WO-US030124.
 XX
 PR 26-SEP-2000; 2000US-0235487P.
 XX
 XX (CEPT-) CEPTYR INC.
 PA
 XX Luche RM, Wei B;
 PI
 DR WPI; 2002-315802/35.
 XX
 PT New DSP-16 polypeptide, useful for identifying modulators of its
 PT activity, which can be used in the treatment of disorders such as
 PT Duchenne muscular dystrophy, or cancer.
 XX
 PS Claim 46; Page; 87pp; English.
 XX
 CC The present invention relates to a new polypeptide, DSP-16, having a 665
 CC amino acid sequence, given in the specification, or a variant having at
 CC least 50 % identical residues, which retains the ability to
 CC dephosphorylate an activated mitogen-activated protein (MAP) kinase. The
 CC invention can be used for identifying agents which modulate DSP-16
 CC activity, for modulation of a proliferative response in a cell, survival
 CC of a cell, or differentiation of a cell. The cell displays contact
 CC inhibition of cell growth or anchorage independent growth and may display
 CC altered intercellular adhesion. The agent may modulate apoptosis, or the
 CC cell cycle. The identified modulators can be used to treat Duchenne
 CC muscular dystrophy, cancer, graft-versus-host disease, autoimmune
 CC diseases, allergies, metabolic diseases, abnormal cell growth, abnormal
 CC cell proliferation, and cell cycle abnormalities. The present amino acid
 CC sequence represents the human dual-specificity phosphatase-3 (DSP-16)
 CC mutant protein #1. Note: This sequence is not shown in the specification
 CC but is derived from the wild-type human DSP-16 (AAU79156) protein given
 CC in figure 2 of the specification
 XX
 SQ Sequence 665 AA;
 Query Match 68.0%; Score 452; DB 5; Length 665;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 452; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 214 SFCEKILPMLDKSVDFIEKAKSNGCVLVHCLAGISRSATIAIAYIMKMDMSLDEAYRF 273
 DB 214 SFCEKILPMLDKSVDFIEKAKSNGCVLVHCLAGISRSATIAIAYIMKMDMSLDEAYRF 273
 QY 274 VKEKRPRTISPNFNLGQLDYEKKIKNOTGASGPKSKLHLKENEVPVAVSEGGQKS 333
 DB 274 VKEKRPRTISPNFNLGQLDYEKKIKNOTGASGPKSKLHLKENEVPVAVSEGGQKS 333
 QY 334 ETPLSPCADSATSEAGQRPVPAVSPVQPSLLSDSPLVQALSGIHLTADRLSDS 393
 DB 334 ETPLSPCADSATSEAGQRPVPAVSPVQPSLLSDSPLVQALSGIHLTADRLSDS 393

Db 334 ETPPLSPDCADSATSEAGRPVHPASVSPVSPQLLEDSPLVQALSGHLNLSADRLSDS 393
 QY 394 NKLKRSFSLDIKSVSASVMAALHGFSSSEDALEYKPTTLDGINKLCQFSPVOELSE 453
 Db 394 NKLKRSFSLDIKSVSASVMAALHGFSSSEDALEYKPTTLDGINKLCQFSPVOELSE 453
 QY 454 QTPETSPPDKEEASIPKKLQTPARPSDSQSKRLHSVRTSSGCTAQRSLSPHRSVSDNY 513
 Db 454 QTPETSPPDKEEASIPKKLQTPARPSDSQSKRLHSVRTSSGCTAQRSLSPHRSVSDNY 513
 QY 514 HNSFLFGLSTSQOHLTKSAGLGLKGMHSDTLAPQTPSTPSTSSWYATSSSHYASALY 573
 Db 514 HNSFLFGLSTSQOHLTKSAGLGLKGMHSDTLAPQTPSTPSTSSWYATSSSHYASALY 573
 QY 574 GGSASYSAYSCQLPTCGDQVSVRRRQKPSDRADRSRSHESPEKQPKRSQCMERG 633
 Db 574 GGSASYSAYSCQLPTCGDQVSVRRRQKPSDRADRSRSHESPEKQPKRSQCMERG 633
 QY 634 ESIMSENRSREELGKVGSSQSSFGSMELIEVS 665
 Db 634 ESIMSENRSREELGKVGSSQSSFGSMELIEVS 665
 RESULT 14
 ID AAU79162 standard; protein; 665 AA.
 XX AAU79162;
 DT 02-JUL-2002 (first entry)
 DE Human dual-specificity phosphatase-3 (DSP-16) mutant protein #2.
 XX Human dual-specificity phosphatase-3; DSP-16; autoimmune disease;
 KM mitogen-activated protein kinase; MAP; Duchenne muscular dystrophy;
 KM cancer; graft-versus-host disease; allergy; metabolic disease;
 KM abnormal cell growth; abnormal cell proliferation; contact inhibition;
 KM cell cycle abnormality; anchorage independent cell growth; apoptosis;
 KM intercellular adhesion; DSP-16 modulator; mutant; mutain.
 XX Homo sapiens.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT Misc-difference 244 /note= "Wild-type Cys substituted by Ser"
 FT WO200226997-A2.
 PN 04-APR-2002.
 PD 25-SEP-2001; 2001WO-US030124.
 XX 26-SEP-2000; 2000US-0235487P.
 PR (CEPT-) CEPTYR INC.
 PA Luche RM, Wei B;
 PI Luche RM, Wei B;
 XX MPI; 2002-315802/35.
 DR New DSP-16 polypeptide, useful for identifying modulators of its
 PT activity, which can be used in the treatment of disorders such as
 PT Duchenne muscular dystrophy, or cancer.
 XX Claim 46; Page; 87pp; English.
 CC The present invention relates to a new polypeptide, DSP-16, having a 665
 CC amino acid sequence, given in the specification, or a variant having at
 CC least 50 % identical residues, which retains the ability to
 CC dephosphorylate an activated mitogen-activated protein (MAP) kinase. The
 CC invention can be used for identifying agents which modulate DSP-16
 CC activity, for modulation of a proliferative response in a cell, survival

CC of a cell, or differentiation of a cell. The cell displays contact
 CC inhibition of cell growth or anchorage independent growth and may display
 CC altered intercellular adhesion. The agent may modulate apoptosis, or the
 CC cell cycle. The identified modulators can be used to treat Duchenne
 CC muscular dystrophy, cancer, graft-versus-host disease, autoimmune
 CC diseases, allergies, metabolic diseases, abnormal cell growth, abnormal
 CC cell proliferation, and cell cycle abnormalities. The present amino acid
 CC sequence represents the human dual-specificity phosphatase-3 (DSP-16)
 CC mutant protein #2. Note: This sequence is not shown in the specification
 CC but is derived from the wild-type human DSP-16 (AAU79156) protein given
 CC in figure 2 of the specification
 XX
 SQ Sequence 665 AA;
 Query Match 63.3%; Score 421; DB 5; Length 665;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 421; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 245 LAGISRSATIAIAYIMKRMDSLDEAYRFVKEKRPISNPNFLGQLDYEKKIKNOTGA 304
 Db 245 LAGISRSATIAIAYIMKRMDSLDEAYRFVKEKRPISNPNFLGQLDYEKKIKNOTGA 304
 QY 305 SGPKSRLKLLHLEKNEPVPAVSEGGQKSETPSPCADSATSEAGRPVHPASVSPVP 364
 Db 305 SGPKSRLKLLHLEKNEPVPAVSEGGQKSETPSPCADSATSEAGRPVHPASVSPVP 364
 QY 365 SVQPSLLEDSPLVQALSGHLNLSADRLSDSNKLKRSFSLDIKSVSASVMAALHGFSSSE 424
 Db 365 SVQPSLLEDSPLVQALSGHLNLSADRLSDSNKLKRSFSLDIKSVSASVMAALHGFSSSE 424
 QY 425 DALEYKPTTLDGINKLCQFSPVOELSEQTPETSPPDKEEASIPKKLQTPARPSDSQSKRL 484
 Db 425 DALEYKPTTLDGINKLCQFSPVOELSEQTPETSPPDKEEASIPKKLQTPARPSDSQSKRL 484
 QY 485 HSVRTSSGCTAQRSLSPHRSVSDNYHTSFLFGLSTSQOHLTKSAGLGLKGMHSDTL 544
 Db 485 HSVRTSSGCTAQRSLSPHRSVSDNYHTSFLFGLSTSQOHLTKSAGLGLKGMHSDTL 544
 QY 545 APQTSPLSTSSWYATSSSHYASALYIGSASYSAYSCQLPTCGDQVSVRRRQKPS 604
 Db 545 APQTSPLSTSSWYATSSSHYASALYIGSASYSAYSCQLPTCGDQVSVRRRQKPS 604
 QY 605 DRADRSRSHESPEKQPKRSQCMERGESIMSENRSREELGKVGSSQSSFGSMELIEV 664
 Db 605 DRADRSRSHESPEKQPKRSQCMERGESIMSENRSREELGKVGSSQSSFGSMELIEV 664
 QY 665 S 665
 Db 665 S 665
 RESULT 15
 ID ABR52352 standard; protein; 665 AA.
 XX ABR52352;
 DT 19-JUN-2003 (first entry)
 DE Protein relating to the invention SEQ ID NO: 42.
 XX
 KM antiproliferative; hepatotropic; nephrotropic; antiarthritic;
 KM antipsoriatic; cardiant; cytostatic; gene therapy; liver disease;
 KM proliferative disorder; renal failure; cardiovascular disorder;
 KM immunological disorder; arthritis; psoriasis; congenital heart defect;
 KM congestive heart failure; cancer; human; mouse; yeast; rat; fruitfly.
 XX Homo sapiens.
 OS
 XX WO200257460-A2.
 PN 25-JUL-2002.
 PD

PF 20-DEC-2001; 2001WO-US050459.
XX
PR 20-DEC-2000; 2000US-0255686P.
PR 30-MAR-2001; 2001US-0280186P.
PR 01-MAY-2001; 2001US-0287735P.
PR 05-JUN-2001; 2001US-0295848P.
PR 25-JUN-2001; 2001US-0300465P.
XX
PA (BRIM) BRISTOL-MYERS SQUIBB CO.
PI Jackson DG, Feder J, Nelson T, Mintier G, Ramanathan C, Lee L,
PI Siemers N, Bol D, Schieven G, Finger J, Todderud CG, Bassolino D,
PI Krystek S, Mcatee P, Suchard S, Banas D;
XX
DR MPI; 2002-599721/64.
DR N-PSDB; ACC60521.
XX
XX Novel polynucleotides encoding human phosphatase polypeptides useful in
PT the prevention or treatment of e.g. proliferative and cardiovascular
PT disorders.
XX
PS Claim 5; Fig 5; 801pp; English.
XX
XX The invention relates to a novel isolated nucleic acid comprising a
CC polynucleotide having a nucleotide sequence selected from 40
CC polynucleotides fully defined in the specification. The polynucleotide of
CC the invention has antiproliferative, hepatotropic, nephrotropic,
CC antiarthritic, antipsoriatic, cardiant, and cytostatic activity. The
CC polynucleotide may have a use in gene therapy. A polynucleotide or
CC polypeptide of the invention is useful for preventing, treating or
CC ameliorating a medical condition, e.g. a proliferative disorder. They are
CC also useful for treating e.g. liver disease, renal failure, immunological
CC disorders including arthritis and psoriasis, cardiovascular disorders
CC such as congenital heart defects and congestive heart failure, and
CC cancer. A method of the invention is useful for diagnosing a pathological
CC condition or susceptibility to a condition in a subject. The present
CC sequence is used in the exemplification of the invention
XX
SQ Sequence 665 AA;
XX
Query Match 52.6%; Score 350; DB 5; Length 665;
Best Local Similarity 100.0%; Pred. No. 2e-317;
Matches 350; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 316 LKKPNEPVAVSEGGQKSETPSPPCADSATSEAAQORVHPVAVSPVPSVPSLLEDS 375
DB 316 LKKPNEPVAVSEGGQKSETPSPPCADSATSEAAQORVHPVAVSPVPSVPSLLEDS 375
QY 376 LVQALSGHLASADRLSDSNKLRKSFSLDIKSVSYASMAASLHGFSSEDALEYKPS 435
DB 376 LVQALSGHLASADRLSDSNKLRKSFSLDIKSVSYASMAASLHGFSSEDALEYKPS 435
QY 436 LDGYNKLCQFSFVQELSEOTPETSPDKERASTPKLQOTARPDSOSKRLHSVRTSSG 495
DB 436 LDGYNKLCQFSFVQELSEOTPETSPDKERASTPKLQOTARPDSOSKRLHSVRTSSG 495
QY 496 QRSLSPLHRSQSVEDNHTSFLFGLSTSQOHLTKSAGLGLGMSHDILAPOTPSPSL 555
DB 496 QRSLSPLHRSQSVEDNHTSFLFGLSTSQOHLTKSAGLGLGMSHDILAPOTPSPSL 555
QY 556 SWYFATESHFSASAIYGSASYSAYSCSOLPTCGDQYVSVRRKQPSDRADSRSM 615
DB 556 SWYFATESHFSASAIYGSASYSAYSCSOLPTCGDQYVSVRRKQPSDRADSRSM 615
QY 616 EBPFEKQFRRSCQMEFGESIMSNNRRELKGVGSSQSSFGSMETIEVS 665
DB 616 EBPFEKQFRRSCQMEFGESIMSNNRRELKGVGSSQSSFGSMETIEVS 665
XX
RESULT 16
ABR52424
ID ABR52424 standard; protein; 664 AA.
XX

AC ABR52424;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein relating to the invention SEQ ID NO: 190.
XX
XX
XX antiproliferative; hepatotropic; nephrotropic; antiarthritic;
XX antipsoriatic; cardiant; cytostatic; gene therapy; liver disease;
XX proliferative disorder; renal failure; cardiovascular disorder;
XX immunological disorder; arthritis; psoriasis; congenital heart defect;
XX congestive heart failure; cancer; human; mouse; yeast; rat; fruitfly.
XX
OS Homo sapiens.
XX
XX WO200257460-A2.
XX
XX 25-JUL-2002.
XX
XX 20-DEC-2001; 2001WO-US050459.
XX
XX
XX 20-DEC-2000; 2000US-0255686P.
XX 30-MAR-2001; 2001US-0280186P.
XX 01-MAY-2001; 2001US-0287735P.
XX 05-JUN-2001; 2001US-0295848P.
XX 25-JUN-2001; 2001US-0300465P.
XX
XX (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
XX Jackson DG, Feder J, Nelson T, Mintier G, Ramanathan C, Lee L,
XX Siemers N, Bol D, Schieven G, Finger J, Todderud CG, Bassolino D,
XX Krystek S, Mcatee P, Suchard S, Banas D;
XX
XX MPI; 2002-599721/64.
XX
XX
XX Novel polynucleotides encoding human phosphatase polypeptides useful in
PT the prevention or treatment of e.g. proliferative and cardiovascular
PT disorders.
XX
XX
XX Example 57; Page 500-501; 801pp; English.
XX
XX The invention relates to a novel isolated nucleic acid comprising a
CC polynucleotide having a nucleotide sequence selected from 40
CC polynucleotides fully defined in the specification. The polynucleotide of
CC the invention has antiproliferative, hepatotropic, nephrotropic,
CC antiarthritic, antipsoriatic, cardiant, and cytostatic activity. The
CC polynucleotide may have a use in gene therapy. A polynucleotide or
CC polypeptide of the invention is useful for preventing, treating or
CC ameliorating a medical condition, e.g. a proliferative disorder. They are
CC also useful for treating e.g. liver disease, renal failure, immunological
CC disorders including arthritis and psoriasis, cardiovascular disorders
CC such as congenital heart defects and congestive heart failure, and
CC cancer. A method of the invention is useful for diagnosing a pathological
CC condition or susceptibility to a condition in a subject. The present
CC sequence is used in the exemplification of the invention
XX
SQ Sequence 664 AA;
XX
Query Match 42.3%; Score 281; DB 5; Length 664;
Best Local Similarity 100.0%; Pred. No. 4.9e-253;
Matches 281; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 303 GASGPKSLKILHLEKPNPVAVSEGGQKSETPSPPCADSATSEAAQORVHPVAVSP 362
DB 302 GASGPKSLKILHLEKPNPVAVSEGGQKSETPSPPCADSATSEAAQORVHPVAVSP 361
QY 363 VPSVQPSLLEDSPLVQALSGHLASADRLSDSNKLRKSFSLDIKSVSYASMAASLHGFS 422
DB 362 VPSVQPSLLEDSPLVQALSGHLASADRLSDSNKLRKSFSLDIKSVSYASMAASLHGFS 421
QY 423 SEDALEYKPSSTTLDGYNKLCQFSFVQELSEOTPETSPDKERASTPKLQOTARPDSOSK 482
DB 422 SEDALEYKPSSTTLDGYNKLCQFSFVQELSEOTPETSPDKERASTPKLQOTARPDSOSK 481
XX

```

QY 483 RLHSVTSSTSSGTAQRLSLPLHRSGSVEDNYHTSLFGLSTSQOHLTKSAGLGLKGMHSD 542
    |||||
DB 482 RLHSVTSSTSSGTAQRLSLPLHRSGSVEDNYHTSLFGLSTSQOHLTKSAGLGLKGMHSD 541
    |||||
QY 543 ILAPQSTPSLTSSWYFATESSHFYASAIYGGASAYSAYS 583
    |||||
DB 542 ILAPQSTPSLTSSWYFATESSHFYASAIYGGASAYSAYS 582
    |||||

RESULT 17
AAB20325
ID AAB20325 standard; protein; 666 AA.
XX
AC AAB20325;
XX
XX 29-MAY-2001 (first entry)
XX
DE Human protein phosphatase and kinase protein-4.
XX
XX Protein phosphatase and kinase protein; PPHK-4; human;
KW gastrointestinal disorder; immune system disorder; neurological disorder;
KW cell proliferative disorder; cancer; diagnosis; therapy.
XX
XX Homo sapiens.
OS
XX
FH Key Location/Qualifiers
FT Modified-site 12 /note= "O-phosphorylated"
FT Modified-site 21 /note= "O-phosphorylated"
FT Modified-site 23 /note= "O-phosphorylated"
FT Modified-site 38 /note= "O-phosphorylated"
FT Modified-site 38 /note= "N-glycosylated"
FT Modified-site 38 /note= "N-glycosylated"
FT Modified-site 49 /note= "N-glycosylated"
FT Modified-site 72 /note= "N-glycosylated"
FT Modified-site 72 /note= "O-phosphorylated"
FT Modified-site 82 /note= "O-phosphorylated"
FT Modified-site 85..298 /note= "O-phosphorylated"
FT Region /note= "VHL-type dual specificity phosphatase signature"
FT Modified-site 91 /note= "O-phosphorylated"
FT Modified-site 190 /note= "O-phosphorylated"
FT Modified-site 212 /note= "N-glycosylated"
FT Modified-site 214 /note= "N-glycosylated"
FT Modified-site 220..280 /note= "O-phosphorylated"
FT Active-site /note= "tyrosine specific protein phosphatase"
FT 237..278
FT Region /note= "Y phosphatase signature"
FT Modified-site 266 /note= "O-phosphorylated"
FT Modified-site 280 /note= "O-phosphorylated"
FT Modified-site 300 /note= "O-phosphorylated"
FT Modified-site 369 /note= "N-glycosylated"
FT Modified-site 369 /note= "O-phosphorylated"
FT Modified-site 393 /note= "O-phosphorylated"
FT Modified-site 421 /note= "O-phosphorylated"
FT Modified-site 422 /note= "O-phosphorylated"
FT Modified-site 434 /note= "O-phosphorylated"
FT Modified-site 439 /note= "O-phosphorylated"

```

```

FT Modified-site /note= "O-phosphorylated"
FT 468 /note= "O-phosphorylated"
FT Modified-site 471 /note= "O-phosphorylated"
FT Modified-site 479 /note= "O-phosphorylated"
FT Modified-site 528 /note= "O-phosphorylated"
FT Modified-site 528 /note= "O-phosphorylated"
FT Modified-site 590 /note= "O-phosphorylated"
FT Modified-site 590 /note= "O-phosphorylated"
FT Modified-site 597 /note= "O-phosphorylated"
FT Modified-site 605 /note= "O-phosphorylated"
FT Modified-site 610 /note= "O-phosphorylated"
FT Modified-site 613 /note= "O-phosphorylated"
FT Modified-site 613 /note= "O-phosphorylated"
FT Modified-site 618 /note= "O-phosphorylated"
FT Modified-site 628 /note= "O-phosphorylated"
FT Modified-site 641 /note= "O-phosphorylated"
FT Modified-site 641 /note= "N-glycosylated"
FT Modified-site 643 /note= "O-phosphorylated"
FT /note= "O-phosphorylated"
XX WO200120004-A2.
XX 22-MAR-2001.
XX PD
XX 14-SEP-2000; 2000WO-US025515.
XX PF
XX 15-SEP-1999; 99US-0154141P.
XX PR
XX (INCY-) INCYTE GENOMICS INC.
XX PA
XX PI Yue H, Tang YT, Bandman O, Hillman JL, Baughn MR, Azimzai Y;
XX PI Lu DM;
XX DR WPI; 2001-244811/25.
XX DR N-PSDB; AAF30479.
XX XX
PT Novel human protein phosphatase and kinase proteins for diagnosis,
PT treatment and prevention of gastrointestinal, immune system, neurological
PT and cell proliferative disorders.
XX PS
XX Claim 1, Page 87-88; 103pp; English.
XX
XX The present sequence is that of novel human protein phosphatase and
XX kinase protein PPHK-4, as predicted from Incyte Clone ID No. 1234795CB1
XX (see AAF30479). Tissues that express PPHK-4 (as a fraction of total
XX tissues expressing PPHK-4) include gastrointestinal (0.385),
XX cardiovascular (0.154), nervous (0.154) and reproductive (0.154).
XX Diseases or conditions associated with tissues expressing PPHK-4 (as a
XX fraction of total tissues expressing PPHK-4) include cancer (0.692),
XX inflammation or trauma (0.308) and cell proliferation (0.231). The
XX encoded protein shows homology to mouse neuronal tyrosine threonine
XX phosphatase 1. The invention provides human PPHK-1 to -11 polypeptides
XX (see AAB20322-32) and polynucleotides (see AAF30476-86). It also provides
XX expression vectors, host cells, antibodies, agonists and antagonists, as
XX well as methods for diagnosing, treating or preventing disorders
XX associated with expression of PPHK, including gastrointestinal
XX disorders, immune system disorders, neurological disorders and cell
XX proliferative disorders, including cancer
XX
XX Sequence 666 AA;

```

Query Match 40.0%; Score 266; DB 4; Length 666;
 Best Local Similarity 100.0%; Pred. No. 4.9e-239;
 Matches 266; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 TCPKPDFPESHFLRPVNDSPCEKILPWLKSVDFIEKAKASNGCVLHCLAGISRAT 253
DB 194 TCPKPDFPESHFLRPVNDSPCEKILPWLKSVDFIEKAKASNGCVLHCLAGISRAT 253
QY 254 IAIAYIMKMDMSLDEAYRFVKEKRPITSPNPNFLGQLLDYEKKIKNOTGASGPKSKL 313
DB 254 IAIAYIMKMDMSLDEAYRFVKEKRPITSPNPNFLGQLLDYEKKIKNOTGASGPKSKL 313
QY 314 LHLKPNPVPVAVSEGGQKSTPLSPPCADATSEAAQRPVHPASVPSVPSLLED 373
DB 314 LHLKPNPVPVAVSEGGQKSTPLSPPCADATSEAAQRPVHPASVPSVPSLLED 373
QY 374 SPLVQALSGHLHSDRLSDSNKLRSPSLDIKSVSYASMAASHGFSSSDALEYKPS 433
DB 374 SPLVQALSGHLHSDRLSDSNKLRSPSLDIKSVSYASMAASHGFSSSDALEYKPS 433
QY 434 TTLDTGNKLCQFSPVQELSEQTPETS 459
DB 434 TTLDTGNKLCQFSPVQELSEQTPETS 459

RESULT 18

AAM25744
ID AAM25744 standard; protein; 672 AA.
XX
AC AAM25744;
XX
DT 16-OCT-2001 (first entry)
XX

Human protein sequence SEQ ID NO:1259.

Human; cancer; HIV infection; human immunodeficiency virus;
anti-inflammatory; antirheumatic; antiarthritic; immunosuppressive;
antibacterial; endocrine; cardiant; central nervous system; virucide;
anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;
antiaggregant; haemostatic; vulnary; antilicer; osteopathic; eczema;
dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;
neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;
immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
genetic disease; haematopoietic disorder; platelet disorder; asthma;
thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
allergic rhinitis; diabetes; multiple sclerosis; depression;
Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
neurological disorder.

XX Homo sapiens.
XX

XX WO200153455-A2.
XX

XX 26-JUL-2001.
XX

XX 22-DEC-2000; 2000WO-US035017.
XX

XX 23-DEC-1999; 99US-00471275.
XX

XX 21-JAN-2000; 2000US-00488725.
XX

XX 25-APR-2000; 2000US-00552317.
XX

XX (HYSB-) HYSEQ INC.
XX

XX Tang YT, Liu C, Drmanac R;
XX

XX WPI, 2001-457603/49.
XX

XX N-PSDB; AAH99685.
XX

XX Isolated human polynucleotides encoding polypeptides, useful for the
XX treatment and diagnosis of e.g. cancer, ulcers and HIV infection.
XX

XX Claim 20; Page 260; 1217p; English.
XX

XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
XX CC AAM25963. The proteins can have activities based on the tissues and cells

CC they are expressed in, such as: anti-inflammatory; antirheumatic;
CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;
CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnary;
CC antilicer; osteopathic; dermatological; antidiabetic; antiasthmatic;
CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC agonists or antagonists of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders

XX SQ Sequence 672 AA;

Query Match 38.8%; Score 258; DB 4; Length 672;
Best Local Similarity 100.0%; Pred. No. 1,4e-211;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 TCPKPDFPESHFLRPVNDSPCEKILPWLKSVDFIEKAKASNGCVLHCLAGISRAT 253
DB 201 TCPKPDFPESHFLRPVNDSPCEKILPWLKSVDFIEKAKASNGCVLHCLAGISRAT 260
QY 254 IAIAYIMKMDMSLDEAYRFVKEKRPITSPNPNFLGQLLDYEKKIKNOTGASGPKSKL 313
DB 261 IAIAYIMKMDMSLDEAYRFVKEKRPITSPNPNFLGQLLDYEKKIKNOTGASGPKSKL 320
QY 314 LHLKPNPVPVAVSEGGQKSTPLSPPCADATSEAAQRPVHPASVPSVPSLLED 373
DB 321 LHLKPNPVPVAVSEGGQKSTPLSPPCADATSEAAQRPVHPASVPSVPSLLED 380
QY 374 SPLVQALSGHLHSDRLSDSNKLRSPSLDIKSVSYASMAASHGFSSSDALEYKPS 433
DB 381 SPLVQALSGHLHSDRLSDSNKLRSPSLDIKSVSYASMAASHGFSSSDALEYKPS 440
QY 434 TTLDTGNKLCQFSPVQEL 451
DB 441 TTLDTGNKLCQFSPVQEL 458

RESULT 19

ABR52385
ID ABR52385 standard; protein; 660 AA.
XX

XX ABR52385;
XX

XX 19-JUN-2003 (first entry)
XX

XX Protein relating to the invention SEQ ID NO: 114.
XX

XX antiproliferative; hepatotropic; nephrotropic; antiarthritic;
XX antiproliferative; cardiant; cytostatic; gene therapy; liver disease;
XX proliferative disorder; renal failure; cardiovascular disorder;
XX immunological disorder; arthritis; psoriasis; congenital heart defect;
XX congestive heart failure; cancer; human; mouse; yeast; rat; fruitfly.

XX Mus musculus.
XX

XX WO200257460-A2.
XX

XX 25-JUL-2002.
XX

XX 20-DEC-2001; 2001WO-US050459.
XX

XX 20-DEC-2000; 2000US-0256868P.
XX

XX 30-MAR-2001; 2001US-0280186P.
XX

PR 01-MAY-2001; 2001US-0287735P.
 PR 05-JUN-2001; 2001US-0295848P.
 PR 25-JUN-2001; 2001US-0300465P.
 XX
 XX
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 PI Jackson DG, Feder J, Nelson T, Minter G, Ramanathan C, Lee L;
 PI Siemsek N, Bol D, Schieven G, Finger J, Todderud CG, Bassolino D;
 PI Krysek S, Mcatee P, Suchard S, Banas D;
 XX WPI; 2002-599721/64.
 DR N-PSDB; ACC60560.
 XX
 PT Novel polynucleotides encoding human phosphatase polypeptides useful in
 PT the prevention or treatment of e.g. proliferative and cardiovascular
 PT disorders.
 PS Disclosure; Fig 12; 801pp; English.
 XX
 CC The invention relates to a novel isolated nucleic acid comprising a
 CC polynucleotide having a nucleotide sequence selected from 40
 CC polynucleotides fully defined in the specification. The polynucleotide of
 CC the invention has antiproliferative, hepatotropic, nephrotropic,
 CC antiarthritic, antipsoriatic, cardiant, and cytostatic activity. The
 CC polynucleotide may have a use in gene therapy. A polynucleotide or
 CC polypeptide of the invention is useful for preventing, treating or
 CC ameliorating a medical condition, e.g. a proliferative disorder. They are
 CC also useful for treating e.g. liver disease, renal failure, immunological
 CC disorders including arthritis and psoriasis, cardiovascular disorders
 CC such as congenital heart defects and congestive heart failure, and
 CC cancer. A method of the invention is useful for diagnosing a pathological
 CC condition or susceptibility to a condition in a subject. The present
 CC sequence is used in the exemplification of the invention
 SQ Sequence 660 AA;
 XX
 Query Match 9.6%; Score 64; DB 5; Length 660;
 Best Local Similarity 100.0%; Pred. NO. 1.4e-50;
 Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 114 SFNSVHLAAGFAFSGRCFGLCEGKSTLVPTCISQPCLEPVANIGTRILPNLYLGGORD 173
 DB 114 SFNSVHLAAGFAFSGRCFGLCEGKSTLVPTCISQPCLEPVANIGTRILPNLYLGGORD 173
 QY 174 VLNK 177
 DB 174 VLNK 177
 XX
 RESULT 20
 ADE08458
 ID ADE08458 standard; protein; 579 AA.
 XX
 AC ADE08458;
 XX
 DT 29-JUN-2004 (first entry)
 XX
 DE Novel protein (useful for identifying genetic disorders) #613.
 XX
 KM novel gene; novel protein; tissue marker; molecular weight marker;
 KM chromosome marker; genetic disorder.
 XX
 OS Unidentified.
 XX
 PN WO2003054152-A2.
 XX
 PD 03-JUL-2003.
 XX
 PF 10-DEC-2002; 2002WO-US039555.
 XX
 PR 10-DEC-2001; 2001US-0339739P.
 PR 11-DEC-2001; 2001US-0339453P.
 PR 14-MAR-2002; 2002US-0365091P.

PR 14-MAR-2002; 2002US-0365384P.
 PR 12-APR-2002; 2002US-0372381P.
 PR 12-APR-2002; 2002US-0372615P.
 PR 22-APR-2002; 2002US-00128558.
 PR 24-APR-2002; 2002US-0376045P.
 XX
 XX (HYSE-) HYSEQ INC.
 XX
 PI Tang XT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;
 PI Ghosh M, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;
 PI Ma Y, Wang D, Chen R, Xu C, Boyle BJ;
 XX WPI; 2003-569235/53.
 DR N-PSDB; ADE07547.
 XX
 PT New polynucleotides, useful for expressing recombinant proteins for
 PT analysis, characterization or therapeutic use, or as markers for tissues
 PT in which the corresponding protein is preferentially expressed.
 PS Claim 20; SEQ ID NO 1524; 1177pp; English.
 XX
 CC The invention comprises the amino acid and coding sequences of novel
 CC proteins. The DNA and protein sequences of the invention are useful as:
 CC markers for tissues in which the corresponding protein is preferentially
 CC expressed; as molecular weight markers on gels; as chromosome markers or
 CC tags; to identify chromosomes or to map related gene positions; and to
 CC compare with endogenous DNA sequences in patients to identify potential
 CC genetic disorders. The present amino acid sequence represents a protein
 CC of the invention.
 SQ Sequence 579 AA;
 XX
 Query Match 3.0%; Score 20; DB 7; Length 579;
 Best Local Similarity 100.0%; Pred. NO. 1.4e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 242 VHCLAGISRSATIAIAYIMK 261
 DB 186 VHCLAGISRSATIAIAYIMK 205
 XX
 RESULT 21
 ABR52382
 ID ABR52382 standard; protein; 625 AA.
 XX
 AC ABR52382;
 XX
 DT 19-JUN-2003 (first entry)
 XX
 DE Protein relating to the invention SEQ ID NO: 110.
 XX
 KM antiproliferative; hepatotropic; nephrotropic; antiarthritic;
 KM antipsoriatic; cardiant; cytostatic; gene therapy; liver disease;
 KM proliferative disorder; renal failure; cardiovascular disorder;
 KM immunological disorder; arthritis; psoriasis; congenital heart defect;
 KM congestive heart failure; cancer; human; mouse; yeast; rat; fruitfly.
 XX
 OS Homo sapiens.
 XX
 PN WO200257460-A2.
 XX
 PD 25-JUL-2002.
 XX
 PF 20-DEC-2001; 2001WO-US050459.
 XX
 PR 20-DEC-2000; 2000US-0256868P.
 PR 30-MAR-2001; 2001US-0280186P.
 PR 01-MAY-2001; 2001US-0287735P.
 PR 05-JUN-2001; 2001US-0295848P.
 PR 25-JUN-2001; 2001US-0300465P.
 XX
 XX (BRIM) BRISTOL-MYERS SQUIBB CO.

PI Jackson DG, Feder J, Nelson T, Mintier G, Ramanathan C, Lee L;
 PI Siemers N, Bol D, Schieven G, Finger J, Todderud CG, Baasolino D;
 PI Krystek S, Mcatee P, Suchard S, Banas D;
 DR WPI; 2002-599721/64.
 XX
 PT Novel polynucleotides encoding human phosphatase polypeptides useful in
 PT the prevention or treatment of e.g. proliferative and cardiovascular
 PT disorders.
 PS Disclosure; Fig 14; 801pp; English.
 CC The invention relates to a novel isolated nucleic acid comprising a
 CC polynucleotide having a nucleotide sequence selected from 40
 CC polynucleotides fully defined in the specification. The polynucleotide of
 CC the invention has antiproliferative, hepatotropic, nephrotropic,
 CC antiarthritic, antipsoptic, cardiant, and cyostatic activity. The
 CC polynucleotide may have a use in gene therapy. A polynucleotide or
 CC polypeptide of the invention is useful for preventing, treating or
 CC ameliorating a medical condition, e.g. a proliferative disorder. They are
 CC also useful for treating e.g. liver disease, renal failure, immunological
 CC disorders including arthritis and psoriasis, cardiovascular disorders
 CC such as congenital heart defects and congestive heart failure, and
 CC cancer. A method of the invention is useful for diagnosing a pathological
 CC condition or susceptibility to a condition in a subject. The present
 CC sequence is used in the exemplification of the invention
 SQ Sequence 625 AA;

Query Match 3.0%; Score 20; DB 5; Length 625;
 Best Local Similarity 100.0%; Pred. No. 1.5e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 242 VHCLAGISRSATIAIAYIMK 261
 |||||
 DB 244 VHCLAGISRSATIAIAYIMK 263

RESULT 22
 ABR52350
 ID ABR52350 standard; protein; 625 AA.
 XX
 AC ABR52350;

DT 19-JUN-2003 (first entry)
 DE Protein relating to the invention SEQ ID NO: 39.

XX antiproliferative; hepatotropic; nephrotropic; antiarthritic;
 XX antipsoptic; cardiant; cyostatic; gene therapy; liver disease;
 KW proliferative disorder; renal failure; cardiovascular disorder;
 KW immunological disorder; arthritis; psoriasis; congenital heart defect;
 KW congestive heart failure; cancer; human; mouse; yeast; rat; fruitfly.
 XX

OS Homo sapiens.
 XX
 PN WO200257460-A2.
 PD 25-JUL-2002.

PF 20-DEC-2001; 2001WO-US050459.
 XX
 PR 20-DEC-2000; 2000US-0256868P.
 PR 30-MAR-2001; 2001US-0280186P.
 PR 01-MAY-2001; 2001US-0287735P.
 PR 05-JUN-2001; 2001US-0295848P.
 PR 25-JUN-2001; 2001US-0300465P.

XX (BRIM) BRISTOL-MYERS SQUIBB CO.
 PA

PI Jackson DG, Feder J, Nelson T, Mintier G, Ramanathan C, Lee L;
 PI Siemers N, Bol D, Schieven G, Finger J, Todderud CG, Baasolino D;
 PI Krystek S, Mcatee P, Suchard S, Banas D;

XX
 DR WPI; 2002-599721/64.
 XX

PT Novel polynucleotides encoding human phosphatase polypeptides useful in
 PT the prevention or treatment of e.g. proliferative and cardiovascular
 PT disorders.
 PS Disclosure; Fig 10; 801pp; English.

CC The invention relates to a novel isolated nucleic acid comprising a
 CC polynucleotide having a nucleotide sequence selected from 40
 CC polynucleotides fully defined in the specification. The polynucleotide of
 CC the invention has antiproliferative, hepatotropic, nephrotropic,
 CC antiarthritic, antipsoptic, cardiant, and cyostatic activity. The
 CC polynucleotide may have a use in gene therapy. A polynucleotide or
 CC polypeptide of the invention is useful for preventing, treating or
 CC ameliorating a medical condition, e.g. a proliferative disorder. They are
 CC also useful for treating e.g. liver disease, renal failure, immunological
 CC disorders including arthritis and psoriasis, cardiovascular disorders
 CC such as congenital heart defects and congestive heart failure, and
 CC cancer. A method of the invention is useful for diagnosing a pathological
 CC condition or susceptibility to a condition in a subject. The present
 CC sequence is used in the exemplification of the invention
 SQ Sequence 625 AA;

Query Match 3.0%; Score 20; DB 5; Length 625;
 Best Local Similarity 100.0%; Pred. No. 1.5e-09;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 242 VHCLAGISRSATIAIAYIMK 261
 |||||
 DB 244 VHCLAGISRSATIAIAYIMK 263

RESULT 23
 AEG73440
 ID AEG73440 standard; protein; 625 AA.
 XX
 AC AEG73440;

DT 10-MAY-2003 (first entry)

DE Human dual specific phosphatase 8 polypeptide.

KW Human; dual specific phosphatase 8; enzyme; infection; inflammation;
 KW tumour formation; cyostatic; antiinflammatory.
 XX

OS Homo sapiens.

PN US6482644-B1.

PD 19-NOV-2002.

PF 01-AUG-2001; 2001US-00920668.

PR 01-AUG-2001; 2001US-00920668.

PA (ISIS-) ISIS PHARM INC.

PI Cowseart LM;

DR WPI; 2003-298140/29.

DR N-PSDB; ABX10760.

PT New antisense compound targeted to a nucleic acid encoding human dual
 PT specific phosphatase 8, for modulating gene expression and treating
 PT diseases associated with expression of the phosphatase in humans.
 XX

PS Disclosure; Col 49-54; 36pp; English.

CC The invention relates to a compound targeted to the coding region of a
 CC nucleic acid encoding human dual specific phosphatase 8, where the

CC compound specifically hybridises with the region and inhibits the
CC expression of human dual specific phosphatase 8. The compound is useful
CC for inhibiting the expression of human dual specific phosphatase 8 in
CC cells or tissues, and for treating an animal, particularly a human,
CC suspected of having or being prone to a disease or condition associated
CC with expression of dual specific phosphatase 8. The compound is useful
CC for diagnostic, therapeutic and as a research reagent, e.g. to prevent
CC or delay infection, inflammation or tumour formation, and to distinguish
CC between functions of various members of a biological pathway. This
CC sequence represents human dual specific phosphatase 8

XX SQ Sequence 625 AA;

Query Match 3.0%; Score 20; DB 6; Length 625;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 242 VHCLAGISRSATTAIAYIMK 261
DB 244 VHCLAGISRSATTAIAYIMK 263

RESULT 24
ID AAW29150 standard; protein; 663 AA.
XX AAW29150;
XX 15-DEC-1997 (first entry)
XX Dual-specific murine threonine-tyrosine phosphatase M3/6.
XX
XX murine; phosphatase; M3/6; dual specificity; Threonine; Tyrosine;
XX inactive; mitogen activated protein kinase; MAP-K; cdc25 PTP; yeast;
XX trinucleotide repeat; repeat expansion; neurodegenerative; proliferative;
XX diagnosis; tumour; lung; brain; chromosomal deletion.
XX
XX Mus sp.
XX
XX Key Location/Qualifiers
XX M3c-difference 29.49 /note="region of homology with yeast cdc25"
XX M3c-difference 117.136 /note="region of homology with yeast cdc25"
XX Domain 244.253 /label=catalytic_domain
XX
XX WO9706245-A1.
XX
XX 20-FEB-1997.
XX
XX 05-AUG-1996; 96WO-G8001906.
XX
XX 04-AUG-1995; 95GB-00016059.
XX
XX (MED1-) MEDICAL RES COUNCIL.
XX
XX Davies KE, Theodosiou A,
XX WPI; 1997-154253/14.
XX N-PDB; AAT86758.
XX
XX Murine mitogen activated protein (MAP) kinase phosphatase M3/6 - is a
XX suspected dual specificity Thr-Tyr phosphatase, useful for diagnosing and
XX treating neuro-degenerative or proliferative diseases e.g. tumours.
XX
XX Claim 23; Fig 2; 51pp; English.

CC This is a murine phosphatase designated M3/6 (encoded by AAT86758). M3/6
CC is a suspected dual specificity Threonine-Tyrosine phosphatase, capable
CC of inactivating mitogen activated protein (MAP) kinase. The M3/6 protein
CC product shows high homology to the cdc25 PTP of yeast at residues 29-49
CC and 117-136. The gene also contains a complex triplet distal to the

CC catalytic domain which is translated into the protein. This domain
CC comprises a run of 4 serine residues which in turn is followed by a
CC further run comprising 23 serine residues which is interrupted near the N
CC terminal section by a single asparagine. This makes the phosphatase gene
CC a candidate for a human disease caused by repeat expansion or mutation.
CC M3/6 is expressed highly in the brain and may have utility in
CC investigating signal transduction mechanisms in brain and muscle. The
CC M3/6 and Hbs (a human homologue) genes may be responsible, if mutated,
CC for various neurodegenerative or proliferative diseases, and may
CC therefore be used for the diagnosis of such diseases, e.g. tumours,
CC especially lung or brain tumours, associated with deletion of the
CC chromosomal region 11p15.5. The polypeptides can be used to screen for
CC inhibitors to treat these diseases

XX SQ Sequence 663 AA;

Query Match 3.0%; Score 20; DB 2; Length 663;
Best Local Similarity 100.0%; Pred. No. 1.6e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 242 VHCLAGISRSATTAIAYIMK 261
DB 244 VHCLAGISRSATTAIAYIMK 263

RESULT 25
ID ABR52351 standard; protein; 663 AA.
XX ABR52351;
XX 19-JUN-2003 (first entry)
XX
XX Protein relating to the invention SEQ ID NO: 40.
XX
XX anti-proliferative; hepatotropic; nephrotropic; antiarthritic;
XX antipsoriatic; cardiac; cytostatic; gene therapy; liver disease;
XX immunological disorder; renal failure; cardiovascular disorder;
XX congestive heart failure; arthritis; psoriasis; congenital heart defect;
XX human; mouse; yeast; rat; fruitfly.
XX
XX Mus musculus.
XX
XX WO200257460-A2.
XX
XX 25-JUL-2002.
XX
XX 20-DEC-2001; 2001WO-US050459.
XX
XX 20-DEC-2000; 2000US-0256868P.
XX
XX 30-MAR-2001; 2001US-0280186P.
XX
XX 01-MAY-2001; 2001US-0287735P.
XX
XX 05-JUN-2001; 2001US-0295848P.
XX
XX 25-JUN-2001; 2001US-0300465P.
XX
XX (BRIM) BRISTOL-MYERS SQUIBB CO.
XX
XX Jackson DG, Feder J, Nelson T, Ramanathan C, Lee L;
XX Stewers N, Bol D, Schieven G, Finger J, Todderud CG, Basolino D;
XX Krysiek S, Mcatee P, Suchard S, Banas D;
XX WPI; 2002-599721/64.
XX
XX Novel polynucleotides encoding human phosphatase polypeptides useful in
XX the prevention or treatment of e.g. proliferative and cardiovascular
XX disorders.
XX
XX Disclosure; Fig 10; 801pp; English.

CC The invention relates to a novel isolated nucleic acid comprising a
CC polynucleotide having a nucleotide sequence selected from 40
CC polynucleotides fully defined in the specification. The polynucleotide of
CC the invention has antiproliferative, hepatotropic, nephrotropic,

CC antiarthritic, antipsoriatic, cardiant, and cytostatic activity. The
CC polynucleotide may have a use in gene therapy. A polynucleotide or
CC polypeptide of the invention is useful for preventing, treating or
CC ameliorating a medical condition, e.g. a proliferative disorder. They are
CC also useful for treating e.g. liver disease, renal failure, immunological
CC disorders including arthritis and psoriasis, cardiovascular disorders
CC such as congenital heart defects and congestive heart failure, and
CC cancer. A method of the invention is useful for diagnosing a pathological
CC condition or susceptibility to a condition in a subject. The present
CC sequence is used in the exemplification of the invention
XX
SQ Sequence 663 AA;

Query Match 3.0%; Score 20; DB 5; Length 663;
Best Local Similarity 100.0%; Pred. No. 1.6e-09;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 242 VHCAGISRSATTAAYIMK 261
Db 244 VHCAGISRSATTAAYIMK 263

RESULT 26

ABB63527
ID ABB63527 standard; protein; 476 AA.

XX ABB63527;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 17373.

XX Drosophila; developmental biology; cell signalling; insecticide;

KM pharmaceutical.

OS Drosophila melanogaster.

PN WO200171042-A2.

PD 27-SEP-2001.

PF 23-MAR-2001; 2001WO-US009231.

PR 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PMD, Myers EW;

DR WPI; 2001-656860/75.

DR N-PSDB; ABL07630.

PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.

XX Disclosure; SEQ ID NO 17373; 21bp + Sequence listing; English.

CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signaling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
CC sequences (ABU16176-ABU30511), and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 476 AA;

Query Match 2.0%; Score 13; DB 4; Length 476;

Best Local Similarity 100.0%; Pred. No. 0.0042;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 246 AGISRSATTAAY 258
Db 218 AGISRSATTAAY 230

RESULT 27

ABG00724
ID ABG00724 standard; protein; 836 AA.

XX ABG00724;

DT 13-FEB-2002 (first entry)

DE Novel human diagnostic protein #715.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

KM food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US008631.

PR 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

XX (HSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR N-PSDB; AAS64911.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

XX Claim 20; SEQ ID NO 31083; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)

CC sequences. (I) is useful as hybridisation probes, polymerase chain

CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,

CC and in recombinant production of (II). The polynucleotides are also used

CC in diagnostics as expressed sequence tags for identifying expressed

CC genes. (I) is useful in gene therapy techniques to restore normal

CC activity of (II) or to treat disease states involving (II). (II) is

CC useful for generating antibodies against it, detecting or quantifying a

CC polypeptide in tissue, as molecular weight markers and as a food

CC supplement. (II) and its binding partners are useful in medical imaging

CC of sites expressing (II). (I) and (II) are useful for treating disorders

CC involving aberrant protein expression or biological activity. The

CC polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic

CC amino acid sequences of the invention. Note: The sequence data for this

CC patent did not appear in the printed specification, but was obtained in

CC electronic format directly from WIPO at

CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 836 AA;

Query Match 1.8%; Score 12; DB 4; Length 836;
Best Local Similarity 100.0%; Pred. No. 0.057;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 240 VLVHCLAGTSRS 251
 |||||
 DB 744 VLVHCLAGTSRS 755

RESULT 28
 ADA44807
 ID ADA44807 standard; protein; 473 AA.
 XX
 AC ADA44807;
 XX
 DT 04-DEC-2003 (first entry)
 XX
 DE CD4/TCR CD3epsilon chain chimeric protein CD4epsilon15, SEQ ID NO:2.
 XX
 KM HIV-1 infection; human immunodeficiency virus-1; CD4+ cell; chimeric CD4;
 KM endoplasmic reticulum; ER retention; envelope protein gp160;
 KM T cell receptor CD3epsilon chain; C-terminal domain; CD4epsilon15;
 KM gene therapy; human; receptor.
 XX
 OS Chimeric.
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Protein 1..458
 FT /label= CD4
 FT Region 459..473
 FT /note= "Part of the C-terminal domain of the T cell
 FT receptor CD3epsilon chain"
 XX
 PN MO2003076468-A1.
 XX
 PD 18-SEP-2003.
 XX
 PF 14-MAR-2003; 2003MO-ES000120.
 XX
 PR 14-MAR-2002; 2002ES-0000616.
 XX
 PA (CNSJ) CONSEJO SUPERIOR INVESTIGACIONES CIENTIF.
 XX
 PI Alarcon Sanchez BJ, San Jose Martinez ME, Zaldivar Notario I;
 XX Gomez Buendia M;
 DR MPI; 2003-779059/73.
 DR N-PSDB; ADA44806.
 XX
 PT Composition for treating or preventing human immune deficiency virus,
 PT comprises CD4 chimeric protein having a protective effect in trans, or
 PT related nucleic acid.
 XX
 PS Claim 5; Page 33-35; 43pp; Spanish.
 XX
 CC The invention relates to a composition for the treatment or prevention of
 CC human immunodeficiency virus-1 (HIV-1) infection. The composition
 CC comprises CD4+ cells that have been transduced with a vector that encodes
 CC a chimeric CD4 molecule which is capable of being retained in the
 CC endoplasmic reticulum (ER). The invention also encompasses the use of a
 CC soluble protein factor produced by CD4+ cells that have been transduced
 CC with a vector encoding a chimeric CD4 protein, and the use of an
 CC expression system encoding a chimeric CD4 protein. The ER-localised
 CC chimeric CD4 molecule binds to the HIV-1 envelope protein gp160,
 CC resulting in HIV-1 retention in the ER and thereby preventing viral
 CC replication. In a specific embodiment, the chimeric CD4 molecule
 CC comprises CD4 fused to 15 amino acids of the C-terminal domain of the T
 CC cell receptor CD3epsilon chain; this chimeric CD4 molecule is designated
 CC CD4epsilon15 (ADA44807). A known chimeric CD4 of similar structure but
 CC containing only 10 amino acids from CD3epsilon can also be used.
 CC Compositions of the invention have an in trans effect on the replication
 CC of HIV-1, and may be used to treat and prevent HIV-1 infection. The
 CC present sequence represents the chimeric CD4 molecule CD4epsilon15, which
 CC is specifically claimed for use in compositions of the invention.

SQ Sequence 473 AA;
 Query Match 1.2%; Score 8; DB 7; Length 473;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 DRADSRRS 612
 |||||
 DB 78 DRADSRRS 85

RESULT 29
 AAY59170
 ID AAY59170 standard; protein; 474 AA.
 XX
 AC AAY59170;
 XX
 DT 14-MAR-2000 (first entry)
 XX
 DE CD4-Ig fusion protein CD4mmu.
 XX
 KM HIV; extracellular; CD4; gp120; immunoglobulin; Ig; fusion protein;
 KM secreted protein; SIV infection; medicament.
 XX
 OS Synthetic.
 OS Homo sapiens.
 XX
 PN CA1340741-C.
 XX
 PD 14-SEP-1999.
 XX
 PF 20-JUN-1989; 89CA-00588749.
 XX
 PR 20-JUN-1989; 89CA-00588749.
 XX
 PA (GEHO) GEN HOSPITAL CORP.
 XX
 PI Seed B;
 DR MPI; 2000-063015/06.
 DR N-PSDB; AA48203.
 XX
 PT New fusion gene encoding immunoglobulin-CD4 fusion proteins, useful in
 PT the treatment of HIV or simian immunodeficiency virus infections.
 XX
 PS Example 1; Page 47-53; 89pp; English.
 XX
 CC The invention provides a fusion gene encoding a fusion protein that
 CC comprises an extracellular CD4 DNA sequence or its fragment which binds
 CC to HIV gp120 when fused to an immunoglobulin (Ig) chain and the DNA
 CC sequence of an Ig heavy or light chain, where the DNA sequence encoding
 CC the variable region has been replaced with the DNA sequence which encodes
 CC extracellular CD4 or its gp120 binding fragment. The fusion protein is
 CC capable of being secreted. The fusion proteins are useful for treating
 CC HIV or SIV infections in animals, preferably humans. They are also useful
 CC for producing medicaments which can be used for treating HIV or SIV
 CC infections in humans. The present sequence represents the fusion protein
 CC CD4mmu where the CD4 is linked to human IgG1 at the Met2 site upstream of
 CC the CH1 region
 XX

SQ Sequence 474 AA;
 Query Match 1.2%; Score 8; DB 3; Length 474;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 DRADSRRS 612
 |||||
 DB 78 DRADSRRS 85

RESULT 30
 AAP93011

ID	AAP93011 standard; protein; 481 AA.
XX	
AC	AAP93011;
XX	
D7	25-MAR-2003 (revised)
DT	03-AUG-1992 (first entry)
XX	
DE	Genetic construct which encodes CD4 linked to human Igm at the Pat site upstream of the CH2 region (fusion protein CD4Pmu).
XX	
KW	Fusion protein; immunoglobulin-like molecule; HIV; SIV; therapy; diagnosis; CD4; gp120; binding fragment; glycoprotein; variable region.
OS	Homo sapiens.
XX	
PX	EPJ25262-A.
PN	
XN	26-JUL-1989.
PD	
PF	20-JAN-1989; 89EP-00100913.
PR	22-JAN-1988; 88US-00147351.
PA	(GENO) GEN HOSPITAL CORP.
PI	Seed B;
DR	WI: 1989-214472/30.
N-P	N-PSTDB; AAN90359.
PT	Immunoglobulin-CD4 fusion proteins - used for treating HIV or SIV infections or detecting HIV or SIV in sample.
PS	Example; Table 4, Page 41-47; 68pp; English.
CC	The fusion protein genes of the invention pref. comprises cDNA sequences which encode CD4 or a fragment which binds gp120 ligated to an expression plasmid which encodes an antibody in which the variable region of the gene has been deleted (see WO87-02671). The CD4 portion of the fusion protein may comprise the complete CD4 sequence, the 370 AA extracellular region and the membrane spanning domain, or the extracellular region. CC heavy chain is pref. from IGM, IgG1 or IgG3. The following are specifically claimed: fusion proteins CD4Hlambdai, CD4Mmu, CD4Pmu, CD4Iambdai, and CD4Mmu (No. 67608), PCDAPlambda (No. 67609) and CCAEIlambda1 (No. 67610). (Updated on 25-MAR-2003 to correct PA field.)
SQ	Sequence 481 AA;
OY	Query Match 1.2%; Score 8; DB 1; Length 481; Best Local Similarity 100.0%; Pred.No. 1.9e+02; Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db	605 DRADSRRS 612 78 DRADSRRS 85
RESULT 31	
AAB19510 ID	AAB19510 standard; protein; 481 AA.
XX AC	
AA	AAB19510;
DT	09-JAN-2001 (first entry)
XX DD	
CDA-IgM fusion protein CH4Pmu.	
CM	CD4; IGM; human; CD4Pmu; fusion protein; immunoglobulin; HIV; SIV; gp120; therapy; diagnosis.
CS	Homo sapiens.
TH Key	Location/Qualifiers

FT	Protein	1..395	
FT		/note="CD4 extracellular region"	
FT	Protein	400..481	
FT		/note="IgM heavy chain partial sequence"	
XX	US6117656-A.		
XX			
PD	12-SEP-2000.		
XX			
PF	07-JUN-1995;	95US-00479353.	
XX			
PR	22-JAN-1988;	88US-00147351.	
PR	23-JAN-1989;	89US-00239596.	
PR	09-JUN-1992;	92US-00896781.	
PR	12-APR-1993;	93US-00057952.	
PR	04-FEB-1994;	94US-00191708.	
PA	(GENO) GEN HOSPITAL CORP.		
XX			
PI	Seed B;		
XX			
DR	WPI, 2000-586558/55.		
XX	N-PSDB; AAA50662.		
PT	CD4-immunoglobulin fusion proteins, useful for targeting gp120 of HIV or		
PT	SIV.		
XX			
PS	Example 1; Col 49-60; 39pp; English.		
XX			
CC	The present sequence is that of fusion protein CD4Pmu comprising the		
CC	extracellular portion of CD4, which binds to HIV gp120, linked at its C-		
CC	terminus to the human IgM heavy chain. To obtain the fusion protein, DNA		
CC	encoding CD4 was linked to IgM DNA at the Pst site upstream of the CM2		
CC	region (see AAA50663). Fusion protein CD4Pmu and a nucleic acid encoding		
CC	it are claimed. Also claimed are a vector comprising the nucleic acid,		
CC	and a method of producing the fusion protein in secreted form using a		
CC	transformed host cell. The fusion protein may further comprise a		
CC	therapeutic agent, radiolabel or NMR imaging agent. The fusion protein		
CC	can be administered to an animal (including humans) for treatment of HIV		
CC	or SIV infection, and can also be used in assays for HIV or SIV, imaging		
CC	and tissue staining. IgM fusion proteins such as CD4Pmu provide complement-		
CC	mediated immunity		
XX			
SQ	Sequence 481 AA;		
Query Match	1.2%; Score 8; DB 3; Length 481;		
Best Local Similarity	100.0%; Pred. No. 1.9e+02;		
Matches	8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
QY	605 DRADSRRS 612		
DB	78 DRADSRRS 85		
RESULT 32			
AAVS1081			
ID	AAVS1081 standard; protein, 481 AA.		
XX			
AC	AAVS1081;		
XX			
DT	23-MAR-2000 (first entry)		
XX			
DE	Human fusion protein CD4Pmu.		
XX			
KW	Fusion protein; human; CD4; IgM; immunoglobulin; gp120;		
KW	anti-human immunodeficiency virus; CD4Pmu.		
XX			
OS	Homo sapiens.		
XX	Synthetic.		
PN	US6004781-A.		
XX			
PD	21-DEC-1999.		

```

XX
PF 04-FEB-1994; 94US-00191708.
XX
PR 22-JAN-1988; 88US-00147351.
PR 23-JAN-1989; 89US-00299596.
PR 09-JUN-1992; 92US-00896781.
PR 12-APR-1993; 93US-00057952.
XX
PA (GEHO ) GEN HOSPITAL CORP.
XX
PI Seed B;
XX
DR WPI; 2000-085792/07.
DR N-PSDB; AA244054.
XX
PT Fusion protein useful for the treatment of human immunodeficiency virus.
XX
PS Example 1; Col 49-58; 39pp; English.
XX
CC This invention describes a novel nucleic acid (I) encoding a fusion
CC protein comprising a DNA sequence encoding amino acids 1-173 of CD4 (II)
CC and a DNA sequence encoding a human immunoglobulin (Ig) heavy or light
CC chain (III). The products of the invention have anti-human
CC immunodeficiency virus (HIV) activity and are capable of binding to
CC gp120. The fusion protein is useful for treating human immunodeficiency
CC virus (HIV) or simian immunodeficiency virus (SIV). This sequence
CC represents the fusion protein CD4Fmu which is constructed from CD4 linked
CC to human Igm upstream of the CH2 region
XX
SQ Sequence 481 AA;
XX
Query Match 1.2%; Score 8; DB 3; Length 481;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 605 DRADSRSS 612
DB 78 DRADSRSS 85
XX
RESULT 33
AAVS9171
ID AAVS9171 standard; protein; 481 AA.
XX
AC AAVS9171;
XX
DT 14-MAR-2000 (first entry)
XX
DE CD4-Ig fusion protein CD4Fmu.
XX
KW HIV; extracellular; CD4; gp120; immunoglobulin; Ig; fusion protein;
KW secreted protein; SIV infection; medicament.
XX
OS Synthetic.
OS Homo sapiens.
XX
PN CA1340741-C.
XX
PD 14-SEP-1999.
XX
PF 20-JAN-1989; 89CA-00588749.
XX
PR 20-JAN-1989; 89CA-00588749.
XX
PA (GEHO ) GEN HOSPITAL CORP.
XX
PI Seed B;
XX
DR WPI; 2000-063015/06.
DR N-PSDB; AA248204.
XX
PT New fusion gene encoding immunoglobulin-CD4 fusion proteins, useful in
PT the treatment of HIV or simian immunodeficiency virus infections.

```

```

XX
PS Example 1; Page 54-60; 89pp; English.
XX
CC The invention provides a fusion gene encoding a fusion protein that
CC comprises an extracellular CD4 DNA sequence or its fragment which binds
CC to HIV gp120 when fused to an immunoglobulin (Ig) chain and the DNA
CC sequence of an Ig heavy or light chain, where the DNA sequence encoding
CC the variable region has been replaced with the DNA sequence which encodes
CC extracellular CD4 or its gp120 binding fragment. The fusion protein is
CC capable of being secreted. The fusion proteins are useful for treating
CC HIV or SIV infections in animals, preferably humans. They are also useful
CC for producing medicaments which can be used for treating HIV or SIV
CC infections in humans. The present sequence represents the fusion protein
CC CD4Fmu where the CD4 is linked to human IgG1 at the Pst site upstream of
CC the CH2 region
XX
SQ Sequence 481 AA;
XX
Query Match 1.2%; Score 8; DB 3; Length 481;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 605 DRADSRSS 612
DB 78 DRADSRSS 85
XX
RESULT 34
AAB00158
ID AAB00158 standard; protein; 507 AA.
XX
AC AAB00158;
XX
DT 08-FEB-2001 (first entry)
XX
DE sCD4-SCFv(17b) HIV single chain antibody fusion protein.
XX
KW Fusion protein; HIV; human immunodeficiency virus; antibody; Fv; AIDS;
KW acquired immune deficiency syndrome; neutralisation; infection;
KW gene therapy; CD4; gp120; glycoprotein; resistance; vaccination;
KW binding domain; single chain antibody; chimera; chimeric protein.
XX
OS Human immunodeficiency virus.
OS Synthetic.
XX
PN WO200055207-A1.
XX
PD 21-SEP-2000.
XX
PF 16-MAR-2000; 2000WO-US006946.
XX
PR 16-MAR-1999; 99US-0124681P.
XX
PA (USSH ) US NAT INST OF HEALTH.
XX
PI Berger EA, Del Castillo CM;
XX
DR WPI; 2000-638183/61.
DR N-PSDB; AA54045.
XX
PT Novel neutralizing bispecific fusion proteins effective in viral such as
PT HIV neutralization, comprises two different binding domains, inducing-
PT binding domain and induced-binding domain functionally linked by linker.
XX
PS Claim 39; Page 46-47; 55pp; English.
XX
CC sCD4-SCFv(17b) is a neutralising bispecific fusion protein capable of
CC binding to two sites of its target protein. The protein comprises a first
CC binding domain capable of binding to an inducing site on the target
CC protein, a second binding domain capable of forming neutralising complex
CC with an induced epitope of the target protein and a linker connecting the
CC binding domains. sCD4-SCFv(17b) comprises a soluble CD4 fragment
CC (containing domains D1 and D2) fused to a single chain Fv portion of

```


CC antibody 17b via a linker. sCD4-SCFv(17b), its variant, analogue or
 CC mimetic is used for inactivating gp120 protein of HIV, and for
 CC neutralising HIV. It is also used for blocking and preventing the binding
 CC of the viral or recombinant gp120 protein to soluble CD4 or lymphocyte
 CC CD4 and for inhibiting HIV replication. The chimeric proteins is
 CC therefore useful for treating HIV infection and also AIDS. It is are
 CC particularly useful in the prevention of infection during or immediately
 CC after HIV exposure (e.g., mother/infant transmission, post-exposure
 CC prophylaxis, and as a topical inhibitor) and for providing long term
 CC resistance to HIV infections and AIDS. Gene therapy is used to secrete
 CC the bispecific protein at mucosal surfaces, such as the vaginal, rectal
 CC or oral mucosa. The fusion proteins is highly potent, broadly cross-
 CC reactive with neutralising antibody with high in vivo activity and no Fc-
 CC mediated undesirable targeting properties. When the fusion protein is
 CC substantially derived from human proteins, it has minimal immunogenicity
 CC and toxicity in humans which is of great value in prevention of infection
 CC during or immediately after HIV exposure

XX SQ Sequence 507 AA;

Query Match 1.2%; Score 8; DB 3; Length 507;
 Best Local Similarity 100.0%; Pred. No. 2e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 DRADSRRS 612
 |||||

DB 78 DRADSRRS 85

RESULT 35

AA20152
 ID AAR20152 standard; protein; 519 AA.

XX AAR20152;

XX 25-MAR-2003 (revised)

DT 31-MAR-1992 (first entry)

XX Human CD4 sequence encoded by PATY.6.

XX Human immunodeficiency virus; HIV; gp 120; AIDS; ARC; glycoprotein;

KW acquired immune deficiency syndrome; AIDS related complex;

XX T helper lymphocytes.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Peptide 1..25

FT /label= signal_sequence

XX WO9118618-A.

XX 12-DEC-1991.

XX 25-MAY-1990; 90US-00529186.

XX 25-MAY-1990; 90US-00529186.

XX (BIOJ) BIOGEN INC.

XX Fisher RA, Hession C, Burkly LC;

XX WPI, 1992-007200/01.

XX N-PSDB; AAQ20327.

XX New immuno-therapeutic human CD4 variants and derivs. - elicit AB

XX production to HIV gp.120, useful in treating, preventing and diagnosing

XX AIDS, ARC and HIV infections.

XX Disclosure; Fig 28; 179pp; English.

XX The sequence was deduced from the DNA sequence of subclone PATY.6, contg.

XX DNA coding for the full-length human CD4. The clone was constructed from

CC plasmids pBG178A and pBG378 (both in US8802940). The DNA can be used to
 CC express recombinant CD4 and analogues for use in diagnosis and treatment
 CC of diseases caused by infective agents whose primary targets are T4+
 CC lymphocytes. See also AAR20148-R20155 and AAR21078. (Updated on 25-MAR-
 CC 2003 to correct PA field.)

XX SQ Sequence 519 AA;

Query Match 1.2%; Score 8; DB 2; Length 519;
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 DRADSRRS 612
 |||||

DB 139 DRADSRRS 146

RESULT 36

AA204703
 ID AAP94703 standard; protein; 524 AA.

XX AAP94703;

XX 25-MAR-2003 (revised)

DT 22-MAR-1991 (first entry)

XX Sequence encoded by T4 lymphocyte cDNA obtained from PBL clone lambda-203

XX -4.

XX HIV; soluble T4; immunotherapeutic; prophylactic; diagnostic; AIDS; ARC.

XX Homo sapiens.

XX Key Location/Qualifiers

FT Misc-difference 67 /note= "AA DESIGNATED NUMBER -23"

FT Misc-difference 90 /note= "AA DESIGNATED NUMBER 1"

FT Misc-difference 92 /note= "MATURE N-TERMINUS"

XX WO8901940-A.

XX 09-MAR-1989.

XX 01-SEP-1988; 88WO-US002940.

XX 04-SEP-1987; 87US-00094322.

XX 07-JAN-1988; 88US-00141649.

XX (BIOJ) BIOGEN INC.

XX Fisher RA, Gilbert W, Sato VL, Flavell RA, Maraganore JM;

XX WPI, 1989-085519/11.

XX N-PSDB; AAN90642.

XX DNA sequences coding for soluble T4-like polypeptide(s) - used in

XX immuno-therapeutic and immunosuppressive compns. and for preventing,

XX treating or detecting AIDS.

XX Disclosure; Fig 3; 207pp; English.

XX The polypeptides encoded are useful in immunotherapeutic, prophylactic

XX and diagnostic compns. They can be used to purify HIV from a sample. The

XX soluble T4 protein-based compns. are useful in treating immunodeficient

XX patients suffering from diseases caused by agents whose primary targets

XX are T4+ lymphocytes. They can be used for preventing, treating or

XX detecting AIDS, ARC and HIV infection. (Updated on 25-MAR-2003 to correct

XX PR field.)

XX Sequence 524 AA;

Query Match	1.2%	Score 8	DB 1	Length 524
Best Local Similarity	100.0%	Pred. No. 2.1e+02		
Matches	8	Conservative	0	Indels 0
		Mismatches	0	Gaps 0
QY	605	DRADSRRS	612	
Db	144	DRADSRRS	151	

RESULT 37
AAR26783
ID AAR26783 standard; protein; 530 AA

DT	24-OCT-2003	(revised)
DT	25-MAR-2003	(revised)
DT	06-FEB-1993	(first entry)
XX		
DE	CD4-IGG2 chimeric heavy chain.	

KM homodimer; soluble CD4; T cell receptor; CD4 antigen; high recovery;
KM chimeric; increased serum half life; HIV infection; AIDS; ss.

05 Homo sapiens.
05 Chimeric.

	Key	Location/Qualifiers
FH	Domain	1. .205
FT		

FT	Domain	206. .302
FT		/label= CH1 domain
FT		

FT	313.	423
FT	/label= hinge domain	
FT	Domain	

	Domain	424 . .530 /label= CH3 domain
FT		
FT		

PN WO9213947-A1.
XX

XX 10-FEB-1992; 92WO-US001143.
PF

08-FEB-1991; 91US-00653684.

VENOG-1 ENOGENICS FILMSTOCK INC.
XX Beaudry GA, Maddon PJ:
PI

WPI; 1992-300034/36.
N-PSDB: AAO28089.

CD4-gamma-2 and CD4-IgG2 chimera(s) and expression vectors - for treatment, prevention and diagnosis of HIV infection.

XX	Claim 15; Fig 4; 90pp; English.
PS	
XY	

CC This sentence represents a CD4-15G2 chimeric heavy chain heterotrimer
CC. It was produced by expression of the coding mutagenised cDNA (produced as
CC described in A4028089) in Dhr-CHO cells. The protein is efficiently
CC assembled intracellularly and effectively secreted from mammalian cells
CC pref. CHO, COS, or myeloma cells as a heterotrimer, enabling high
CC recovery and purification from the medium of cells expressing it. It
CC possesses increased serum half-life and has increased avidity for HIV cf
CC heavy chain dimers. It can inhibit HIV infection of CD4+ cells and block
CC the spread of HIV infection within a patient. Attachment to a detectable
CC marker makes it useful in an assay for HIV or SIV infection, and it can
CC also be linked to toxins, eg Diphtheria, Pseudomonas exotoxin A (domains
CC I or II) or the deglycosylated A-chain of ricin. (Updated on 25-MAR-2003
CC to correct PN field.) (Updated on24-OCT-2003 to standardise OS field)

```

SQ Sequence 530 AA;
Query Match Similarity 1.2%; Score 8; DB 2; Length 530;
Best Local Similarity 100.0%; Freq. No. 2.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 605 DRADSRRS 612
    |||||
    |||||
db 78 DRADSRRS 85

```

RESULT 38
AAR46679
ID AAR46679 standard; protein; 530 AA.

AC	ARR46679;	
XX		
DT	25-MAR-2003	(revised)
DT	08-AUG-1994	(first entry)

CD4-IgG2 chimeric heavy chain.

KM CD4; gamma; heavy chain; chimeric; chimaeric; immunoconjugate; HIV;
KM human immunodeficiency virus; radionuclide; toxin; therapy; treatment;
KM imaging; detection; targeting; immunoglobulin; IgG.

OS Homo sapiens.

FH	Key	Location/Qualifiers
FT	Region	1. .204

FT	Region	205. .302
FT		/label= CH1 Region.

FT	/label= Hinge Region.
FT	315. .423
Region	

FT	Region	424. .530
FT		/label= CH3 Region.

PN WO9403191-A1.

PD 17-FEB-1994.
XX

XX 07-AUG-1992; 92US-00927931.
PR

PA (PROG-) PROGENICS PHARM INC.
XX

XX
DR WPI; 1994-065392/08.

	Non-peptidyl toxin or radionuclide and CD4-gamma 2 or CD4-IgG2
XX	
PT	

PT stage HIV infection.

PS Disclosure; Fig 4; 142pp; English.

A tetramer comprising 2 IgG3 heavy chains or two CD4-IgG2 chimeric heavy chains and two kappa light chains or C κ -kappa light chains (AAR46680) linked to a non-peptidyl toxin or a gamma radiation-emitting radionuclide CC of low to moderate cytotoxicity. The resulting immunconjugate comprising CC the toxin can be used to kill HIV infected cells and to treat HIV CC infected subjects to reduce the population of HIV infected cells. It can CC also be used to reduce the likelihood of infection. The immunconjugate CC comprising the radionuclide can be used to image HIV infected tissue, to CC calculate the stage of HIV infection or the efficacy of an anti-HIV CC treatment using the imaging technique and for determining the prognosis CC of an HIV infected subject. (Updated on 25-MAR-2003 to correct PN field.) CC

Sequence 530 AA;
Query Match 1.2%; Score 8; DB 2; Length 530;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 605 DRADSRSS 612
|||
Db 78 DRADSRSS 85

RESULT 39
AAV85080
ID AAV85080 standard; protein; 530 AA.
XX
AC AAV85080;
XX
DT 19-JUN-2000 (first entry)
XX
DE CD4-IgG2 chimeric heterotetramer heavy chain amino acid sequence.
XX
KM CD4-IgG2 chimeric heavy chain heterotetramer; immunocjugate; treatment;
KM cytotoxic radionuclide; cell surface glycoprotein; prevent; infection;
KM cellular immune response interaction mediator; HIV interaction; staging;
KM prognosis; envelope glycoprotein burden; human.
XX
OS Homo sapiens.
XX
PN US6034223-A.
XX
PD 07-MAR-2000.
XX
PF 07-JUN-1995; 95US-00477460.
XX
PR 07-AUG-1992; 92US-00927931.
PR 06-AUG-1993; 93WO-US007422.
PR 03-FEB-1995; 95US-00379516.
XX
PA (PROG-) PROGENICS PHARM INC.
XX
PI Allaway GP, Maddon PJ;
XX
DR WPI; 2000-269502/23.
DR N-PSDB; AA298856.
XX
PT New immunocjugate, used to treat, prevent or image human immune
PT deficiency virus infection, comprises radionuclide attached to
PT heterotetramer of CD4-immunoglobulin chimeras.
XX
PS Disclosure; Fig 4; 58pp; English.
XX
CC This sequence represents the CD4-IgG2 chimeric heavy chain amino acid
CC sequence from the CD4-IgG2 chimeric heterotetramer. The invention relates
CC to an immunocjugate comprising a cytotoxic radionuclide and a
CC heterotetramer of two heavy chains and two light chains. The cytotoxic
CC radionuclide is linked to either the heavy chains or the light chains, or
CC to all four chains, directly or through a bifunctional chelator. Both
CC heavy chains are chimeric CD4-Ig (immunoglobulin) G2 chains encoded by
CC vector CD4-IgG2HC-PRCCMV (ATCC 75193) and both light chains are chimeric
CC CD4-kappa chains encoded by vector CD4-KLC-PRCCMV (ATCC 75194). CD4 is a
CC non-polyomorph cell surface glycoprotein that is expressed on the
CC surface of helper T lymphocytes, cells of the monocyte/macrophage lineage
CC and dendritic cells. CD4 associates with major histocompatibility complex
CC (MHC) class II molecules on the surface of antigen presenting cells to
CC mediate efficient cellular immune response interactions. In humans CD4 is
CC the target of interaction with the human immunodeficiency virus HIV. The
CC immunocjugate is used to kill cells infected with HIV, and for treating
CC or preventing infection. It is also used for imaging HIV-infected tissues
CC (for staging or prognosis of infection, and for assessing efficacy of
CC treatments). The immunocjugate is also used to determine the HIV
CC envelope glycoprotein burden, once determined, this information is used
CC in the staging and prognosis of HIV infected patients. The
CC immunocjugate should be active against all strains of HIV (since the

CD4-gp120 interaction is essential for infection). The heterotetramers
CC are assembled intracellularly and secreted efficiently from mammalian
CC cells, allowing high recovery and purification from the culture medium.
CC They have longer half-life in serum and greater avidity than heavy chain
CC dimers
XX
SQ Sequence 530 AA;
Query Match 1.2%; Score 8; DB 3; Length 530;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 605 DRADSRSS 612
|||
Db 78 DRADSRSS 85

RESULT 40
AAB67323
ID AAB67323 standard; protein; 530 AA.
XX
AC AAB67323;
XX
DT 23-APR-2001 (first entry)
XX
DE CD4-IgG2 chimeric heavy chain protein.
XX
KM Immunocjugate; chelator; chimeric; HIV; human immunodeficiency virus.
KM Homo sapiens.
XX
OS
XX
PN US6177549-B1.
XX
PD 23-JAN-2001.
XX
PF 10-JUN-1999; 99US-00329916.
XX
PR 07-AUG-1992; 92US-00927931.
PR 06-AUG-1993; 93WO-US007422.
PR 03-FEB-1995; 95US-00379516.
PR 07-JUN-1995; 95US-00477460.
XX
PA (PROG-) PROGENICS PHARM INC.
XX
PI Maddon PJ, Allaway GP;
XX
DR WPI; 2001-158582/16.
XX
PT Immunocjugate for treating human immunodeficiency virus-infected
PT subject, consists of cytotoxic radionuclide linked to heterotetramer
PT comprising two chimeric CD4-IgG2 heavy chains and two chimeric CD4-
PT kappa/light chains.
XX
PS Disclosure; Fig 4; 43pp; English.
XX
CC The present invention relates to an immunocjugate, comprising a
CC cytotoxic radionuclide linked, directly or via a bifunctional chelator,
CC to a heterotetramer having two chimeric CD4-IgG2 heavy chains encoded by
CC an expression vector CD4-IgG2HC-PRCCMV and two chimeric CD4-kappa light
CC chains encoded by an expression vector CD4-KLC-PRCCMV. The invention is
CC useful for killing human immunodeficiency virus (HIV)-infected cells, for
CC the treatment and prevention of infection with HIV
XX
SQ Sequence 530 AA;
Query Match 1.2%; Score 8; DB 4; Length 530;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 605 DRADSRSS 612
|||
Db 78 DRADSRSS 85

```

RESULT 41
AAB80884
ID AAB80884 standard; protein; 530 AA.
XX
XX AAB80884;
XX
XX 29-MAY-2001 (first entry)
XX
XX Human CD4-IgG2 chimeric heavy chain.
XX
XX Human; Anti-HIV; CD4-IgG2 chimeric heterotetramer;
XX immunoglobulin gamma 2.
XX
XX Homo sapiens.
XX
XX US6187748-B1.
XX
XX 13-FEB-2001.
XX
XX 07-JUN-1995; 95US-00485372.
XX
XX 08-FEB-1991; 91US-00653684.
XX
XX 10-FEB-1992; 92WO-US001143.
XX
XX 08-DEC-1992; 92US-00960440.
XX
XX (PROG-) PROGENICS PHARM INC.
XX
XX Maddon PJ, Beaudry GA;
XX
XX WPI; 2001-264981/27.
XX
XX N-PSDB; AAF77830.
XX
XX Inhibiting human immunodeficiency virus (HIV) infection of a CD4+ cell,
XX or erecting a subject having CD4+ cells infected with HIV involves using
XX CD4-IgG2 chimeric heterotetramer to form a complex with the HIV.
XX
XX Disclosure; Fig 4; 55pp; English.
XX
XX The present invention relates to a method for inhibiting infection of a
XX CD4+ cell by HIV. The method comprises contacting the HIV with a CD4-IgG2
XX chimeric heterotetramer to form a complex with the HIV (CD4 = cluster of
XX differentiation 4; IgG2 = immunoglobulin gamma 2). CD4 is a cell surface
XX glycoprotein that is expressed primarily on the surface of T cells. In
XX man, CD4 is the target of interaction with HIV. The heterotetramer has
XX two heavy and two light chains which are encoded by expression vectors
XX CD4-IgG2HC-pRCMV (VI) and CD4-KLC-pRCMV (V2), respectively. The method
XX is used to inhibit infection of a CD4+ cell by a HIV and to prevent CD4+
XX cells of a subject from becoming infected with HIV. The method is also
XX useful for treating a subject having CD4+ cells infected with HIV. The
XX present sequence is human fusion protein: CD4-Ig2 chimeric heavy chain of
XX the CD4-IgG2 chimeric heterotetramer. This sequence was used in the
XX method of the present invention
XX
XX Sequence 530 AA;
XX
XX Query Match 1.2%; Score 8; DB 4; Length 530;
XX Beatt Local Similarity 100.0%; Pred.No. 2.1e+02;
XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0
XX
XX 605 DRADSRSS 612
XX |||||
XX 78 DRADSRSS 85
XX
XX RESULT 42
XX ABB71123
XX ID ABB71123 standard; protein; 530 AA.
XX
XX ABB71123;
XX
XX 17-JAN-2003 (first entry)
XX
XX

```

DE	CD4-immunoglobulin G2 (IgG2) chimeric heterotetramer.
XX	
KM	CD4, immunoglobulin G2; Ig gamma2; human immunodeficiency virus-1; HIV-1;
KW	mutant; mutein.
OS	Homo sapiens.
OS	Synthetic.
XX	
PH	Key
FT	Peptide
FT	1. .25
FT	/label= signal_peptide
FT	26. .530
FT	/note= "Mature CD4-IgG2 chimeric heterotetramer"
XX	
XX	US6451313-B1.
PN	
PD	17-SEP-2002.
XX	
PF	07-JUN-1995; 95US-00484681.
XX	
PR	08-FEB-1991; 91US-00653684.
PR	10-FEB-1992; 92WO-US001143.
PR	08-DEC-1992; 92US-00960440.
XX	
PA	(PROG-) PROGENICS PHARM INC.
XX	
PI	Maddon PJ, Beaudry GA;
XX	
DR	WPI; 2003-038273/03.
DR	N-PsDB; ABS55721.
XX	
PT	Novel CD4-immunoglobulin G2 chimeric heterotetramer neutralizes human
PT	immunodeficiency virus-1 with two heavy and light chains encoded by
PT	expression vectors designated CD4-IgG2HC-prcCMV and CD4-KLC-prcCMV,
PT	respectively.
XX	
PS	Claim 1; Fig 4A-H; 54pp; English.
XX	
CC	The invention describes a purified CD4-immunoglobulin (Ig)G2 chimeric
CC	heterotetramer (I) that neutralises human immunodeficiency virus-1 (HIV-
CC	1) having two heavy chains encoded by an expression vector designated CD4-
CC	-IgG2HC-prcCMV, and two light chains encoded by expression vector
CC	designated CD4-KLC-prcCMV. (I) and a composition (II) comprising (I) or
CC	(I) linked to a toxin, are useful for inhibiting HIV infection of a CD4
CC	cell, and preventing a subject being infected with HIV by blocking the
CC	spread of HIV infection. This is the amino acid sequence of the CD4-
CC	immunoglobulin G2 chimeric heavy chain chimeric heterotetramer useful in
CC	inhibiting HIV infection
XX	
SO	Sequence 530 AA;
XX	
Query Match	1.2%; Score 8; DB 6; Length 530;
Best Local Similarity	100.0%; Pred. No. 2.1e+02;
Matches	8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	605 DRADSRRS 612
DB	78 DRADSRRS 85
RESULT 43	
AAR27278	
ID	AAR27278 standard; protein; 532 AA.
XX	
AC	AAR27278;
XX	
DT	25-MAR-2003 (revised)
DT	28-JUL-1995 (first entry)
XX	
DE	CD4:gamma peptide chimeric protein.
XX	
KM	Fusion protein; CD4; extracellular domain; zeta; eta; gamma;
KW	membrane spanning domain; intracellular domain; type I;

```

KM integral membrane homodimer; TCR; T cell antigen receptor;
KW extracellular domain; mouse; human; receptor; chimera;
XX HPB-ALu tumour cell line; natural killer cell.
OS Homo sapiens.
XX MO9215322-AL.
XX PN
XX PD
XX 17-SEP-1992.
XX PF 06-MAR-1992; 92MO-US001785.
XX PR 07-MAR-1991; 91US-00665961.
XX PA (GEHO ) GEN HOSPITAL CORP.
XX PI Seed B, Romeo C, Kolanus W;
XX PS
XX DR WPI; 1992-331474/40.
XX DR N-PSDB; AAQ28706.
XX PT Therapeutic cells expressing chimeric receptors - directing cellular
XX PT response to an infective agent, useful in treating HIV-1, AIDS
XX PT Pneumocystis carinii infections etc.
XX PS
XX SQ Example 2; Page 74-76; 114pp; English.
XX
CC This sequence represents a fusion protein between the CD4 extracellular
CC domain and the gamma protein membrane spanning domain and intracellular
CC domain. The Fc-receptor-associated gamma chain is expressed in cell
CC surface complexes with additional polypeptides, some of which mediate
CC ligand recognition, and others which have undefined function. Gamma bears
CC a homodimeric structure and overall organisation very similar to that of
CC zeta (see also AAQ28704), and is a component of both the mast
CC cell/basophil high affinity Igb receptor, Fc-epsilon-RI, which consists
CC of at least three distinct polypeptide chains and one of the low affinity
CC receptors for IgG, represented in mice by Fc-gamma-RII-alpha. In the
CC production of the CD4 receptor chimera, the gamma cDNA was isolated from
CC the HPB-ALu tumour cell line and from human natural killer cells. The
CC gamma cDNA was joined to the extracellular domain by engineering a BamHI
CC site just upstream of the membrane spanning domain, by a BamHI site
CC naturally present a few residues upstream of the membrane spanning
CC domain. (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ Sequence 532 AA;

```

Query Match 1.2%; Score 8; DB 2; Length 532;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 605 DRADSRRS 612
   |||||
Db 78 DRADSRRS 85

RESULT 44
ID AAR78678 standard; protein; 532 AA.
XX AAR78678;
XX AC
XX AAR78678;
XX DT 16-APR-1996 (first entry)
XX DE T-cell receptor eta.
XX
XX Chimeric receptor; CD4; T-cell receptor eta; HIV; cytolysis;
KW human immunodeficiency virus; adoptive immunotherapy.
XX
XX Homo sapiens.
XX OS
XX MO9521528-AL.
XX PN
XX PD 17-AUG-1995.

```

```

XX
XX 12-JAN-1995; 95MO-US000454.
XX PF 14-FEB-1994; 94US-00195395.
XX PR 02-AUG-1994; 94US-00284391.
XX
XX (GEHO ) GEN HOSPITAL CORP.
XX PA
XX PI Seed B, Banapour B, Romeo C, Kolanus W;
XX PS
XX DR WPI; 1995-292893/38.
XX DR N-PSDB; AAQ96124.
XX
XX Target cytolysis of HIV-infected cells - by chimeric CD4 receptor-bearing
XX PT cells.
XX PT
XX PS
XX PS Example 2; Page 78-79; 118pp; English.
XX
CC Fusion proteins comprising the extracellular domain of CD4 fused to T-
CC cell receptor zeta, gamma or eta (AAR78676-78, respectively) were
CC expressed in CV1 using a vaccine virus vector. These CD4:zeta, CD4:gamma
CC and CD4:eta chimeric receptors mediated cytolysis of targets expressing
CC HIV gp120/41
XX
SQ Sequence 532 AA;

```

Query Match 1.2%; Score 8; DB 2; Length 532;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 605 DRADSRRS 612
   |||||
Db 78 DRADSRRS 85

RESULT 45
ID AAR89458 standard; protein; 532 AA.
XX AAR89458;
XX AC
XX AAR89458;
XX DT 26-SEP-1996 (first entry)
XX DE CD4:eta fusion protein.
XX
XX CD7; transmembrane domain; chimeric receptor; CD5; CD34; CH2; CH3; IgG1;
KW human; CD4; HIV; proteinaceous alpha-helix; T cell; B cell; neutrophil;
KW dendritic cell; therapy; mammal; infection.
XX
XX Synthetic.
XX OS
XX MO9603883-AL.
XX PN
XX PD 15-FEB-1996.
XX
XX 26-JUL-1995; 95MO-US009468.
XX PF
XX PR 02-AUG-1994; 94US-00284391.
XX PR 24-FEB-1995; 95US-00394388.
XX
XX (GEHO ) GEN HOSPITAL CORP.
XX PA
XX PI Seed B, Banapour B, Romeo C, Kolanus W;
XX PS
XX DR WPI; 1996-129034/13.
XX DR N-PSDB; AAT10803.
XX
XX Membrane-bound chimeric receptor comprising extracellular portion
XX including CD4 fragment - cells expressing receptor can be used for
XX treatment of HIV infection.
XX
XX Example 2; Page 80-81; 134pp; English.

```

CC AAT10801-T10803 represent membrane bound proteinaceous chimeric receptors
 CC of the invention. This sequence represents the CD4:eta chimera. The
 CC transmembrane region of the chimeric receptor acts to separate the
 CC intracellular and extracellular domains of the chimera, and contains a
 CC portion of the CD7 (see AAR89440), CD5 or CD4 transmembrane domains.
 CC Alternatively, the extracellular portion of the receptor can be separated
 CC from the intracellular domain by the hinge, CH2 and CH3 domains of human
 CC IgG1 (see AAR89441). The extracellular portion of the chimeric receptor
 CC contains a fragment of CD4 (amino acids 1-394 or 1-200 of the CD4
 CC sequence, see AAR89450 and AAR89451) which specifically recognises and
 CC binds HIV-infected cells, but does not mediate HIV infection. The
 CC extracellular domain of the receptor is separated from the cell membrane
 CC by 48 or 72 angstroms, or by one or more proteinaceous alpha-helices. The
 CC cells expressing the receptor are preferably T cells, B cells,
 CC neutrophils, or dendritic cells. The therapeutic cells expressing the
 CC chimeric receptor are administered to a mammal to treat HIV infection

XX Sequence 532 AA;
 SQ

Query Match 1.2%; Score 8; DB 2; Length 532;
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 DRADSRRS 612
 |||||
 78 DRADSRRS 85

Db

RESULT 46
 AAW02215
 ID AAW02215 standard; protein; 532 AA.
 XX
 AC AAW02215;
 XX
 DT 16-OCT-2003 (revised)
 DT 11-NOV-1996 (first entry)
 XX
 DE CD4:T-cell receptor eta chain chimeric receptor.
 XX
 KW Chimeric receptor; cellular immunity; adoptive immunotherapy; CD4;
 KW human immunodeficiency virus type 1; HIV-1; AIDS; therapy;
 KW T-cell receptor eta chain; cytotoxic T lymphocyte; CTL.
 XX
 OS Homo: sapiens.
 OS Mus sp.
 OS Chimeric.
 OS
 FH Key Location/Qualifiers
 FT Domain 1..393
 FT /label= "Extracellular domain"
 FT /note= "CD4 extracellular domain"
 FT 394..396
 FT /label= Linker
 FT /note= "encoding DNA contains a BamHI site used for
 FT fusion construction"
 FT 397..532
 FT /note= "region of fusion derived from eta chain,
 FT preferred signal-transducing portions for constructs of
 FT the invention are amino acids 423-455, 438-455,
 FT 461-494, 494-528 or 400-420"
 FT 400..437
 FT /label= "Transmembrane domain"
 FT /note= "eta chain transmembrane domain"
 FT 438..575
 FT /label= "intracellular domain"
 FT /note= "eta chain intracellular domain"
 XX
 PN W09625953-A1.
 XX
 PD 29-AUG-1996.
 XX
 PF 25-JAN-1996; 96MO-US001056.
 XX

PR 24-FEB-1995; 95US-00394176.
 XX
 PA (GEHO) GEN HOSPITAL CORP.
 XX
 PI Seed B, Romeo C, Kolanus W;
 XX
 DR WPI; 1996-402134/40.
 DR N-PSDB; AAT36760.
 XX
 PT Direction of cellular immune response using therapeutic cell expressing 2
 PT chimeric receptors - comprising region binding to target cell and region
 PT that signals target cell destruction, or CD28 region, partic. for
 XX eliminating HIV-infected cells.
 PS Claim 7; Page 77-78; 120pp; English.
 XX
 CC A chimeric receptor (AAW0215) comprises the extracellular domain of an
 CC engineered form of the CD4 cellular receptor for HIV and the
 CC transmembrane and intracellular regions, including the cytoytic signal-
 CC transducing portion, of the mouse T-cell receptor eta chain. It can be
 CC obtd. by inserting a gene fusion (AAT36760) into a vaccinia virus vector
 CC and expression in host cells. Chimeric receptors comprising CD4 fused to
 CC eta, eta (see also AAW02213) or Fc receptor gamma (see also AAW02213)
 CC chains are capable of directing cytotoxic T lymphocytes to specifically
 CC recognise and kill cells expressing HIV gp120, thus providing a therapy
 CC for AIDS. (Updated on 16-OCT-2003 to standardise OS field)

XX Sequence 532 AA;
 SQ

Query Match 1.2%; Score 8; DB 2; Length 532;
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 DRADSRRS 612
 |||||
 78 DRADSRRS 85

Db

RESULT 47
 AAW83141
 ID AAW83141 standard; protein; 532 AA.
 XX
 AC AAW83141;
 XX
 DT 03-FEB-1999 (first entry)
 DT
 XX
 DE Chimeric receptor containing human eta polypeptide.
 XX
 KW Human; zeta; eta; gamma; membrane-bound chimeric receptor; infection;
 KW tumour; cancer cell; autoimmunity-generated cell; T cell receptor; CD3;
 KW CD4; B cell receptor; Fc receptor; pathogen; bacterial; fungal;
 KW protozoan; viral.
 XX
 OS Synthetic.
 OS Homo sapiens.
 OS
 FH Key Location/Qualifiers
 FT Domain 1..393
 FT /label= "Extracellular domain"
 FT /note= "CD4 extracellular domain"
 FT 394..396
 FT /label= Linker
 FT /note= "encoding DNA contains a BamHI site used for
 FT fusion construction"
 FT 397..532
 FT /note= "region of fusion derived from eta chain,
 FT preferred signal-transducing portions for constructs of
 FT the invention are amino acids 423-455, 438-455,
 FT 461-494, 494-528 or 400-420"
 FT 400..437
 FT /label= "Transmembrane domain"
 FT /note= "eta chain transmembrane domain"
 FT 438..575
 FT /label= "intracellular domain"
 FT /note= "eta chain intracellular domain"
 XX
 PN (GEHO) GEN HOSPITAL CORP.
 XX
 PD Romeo C, Kolanus W, Seed B;
 XX
 PF WPI; 1999-044582/04.
 DR N-PSDB; AAV70157.
 XX

FT Membrane-bound chimeric receptors - comprising extracellular portion
 PT which recognises and binds a target cell and an intracellular portion of
 PT e.g. a T-cell receptor.
 XX
 PS Claim 11, Col 45-48; 57pp; English.
 CC The present invention describes DNA encoding a membrane-bound chimeric
 CC receptor comprising: (a) an extracellular portion that specifically
 CC recognises and binds a target cell or a target infective agent; and (b)
 CC an intracellular portion of a T-cell receptor CD3, zeta or eta
 CC polypeptide, a B-cell receptor polypeptide or an Fc receptor polypeptide.
 CC The present sequence represents a chimeric receptor containing the human
 CC eta polypeptide. Cells expressing chimeric receptors of the present
 CC invention can be administered to mammals in order to destroy pathogens
 CC (e.g. bacteria, fungi, protozoa or viruses, especially HIV), cancer cells
 CC or autoimmune-generated cells
 CC
 SQ Sequence 532 AA;
 QY Query Match 1.2%; Score 8; DB 2; Length 532;
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 605 DRADSRSS 612
 78 DRADSRSS 85
 RESULT 48
 AAR26531
 ID AAR26531 standard; protein; 534 AA.
 AC AAR26531;
 XX
 XX 25-MAR-2003 (revised)
 DT 28-JAN-1993 (first entry)
 XX
 DE Sequence of CD4-IgG1 chimeric heavy chain heterotetramer.
 XX
 XX CD4-gamma 1 chimeric heavy chain homodimer; expression vector; HIV;
 KM therapy; diagnostic agent; inhibition.
 KW
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Region 205..302
 FT /label=CH1
 FT 303..317
 FT Region /label=hinge
 FT 318..427
 FT /label=CH2
 FT 428..534
 FT Region /label=CH3
 FT
 FT
 PN WO9213559-A1.
 XX
 PD 20-AUG-1992.
 XX
 PF 10-FEB-1992; 92WO-US001152.
 XX
 PR 08-FEB-1991; 91US-00654205.
 XX
 PA (PROG-) PROGENICS PHARM INC.
 XX
 PI Beaudry GA, Maddon PJ;
 XX
 DR WPI; 1992-299758/36.
 DR N-PsDB; AaQ27831.
 XX
 PT CD4-gamma 1 chimeric heavy chain homo-dimer and its expression vector -
 PT for preventing and treating HIV infection useful as a diagnostic agent.
 XX
 PS Example; Fig 4; 88pp; English.

XX The human CD4 cDNA is excised from the plasmid pSPeT4 and cloned into
 CC M3mp18. In order to excise a fragment containing the CH1 exon of the
 CC human gamma 1 heavy chain gene, the plasmid pBR gamma 1 is digested with
 CC SacII, and the SacII sites are then made flush using T4 DNA polymerase.
 CC The fragment containing the CH1 exon is then purified and ligated to the
 CC M3mp18(Cd4) vector. Oligonucleotide-mediated site-directed mutagenesis
 CC is then performed to juxtapose the CD4 and CH1 sequences in frame. The
 CC CD4-CH1 chimeric gene is then linearized and ligated to the pSL1-Pac1 DNA
 CC fragment of the plasmid pBR gamma 1 containing the hinge, CH2, and CH3
 CC exons of the human gamma 1 chain gene designated CD4-IgG1HC-pRCMV (ATCC
 CC 75192). (Updated on 25-MAR-2003 to correct PN field.)
 CC
 SQ Sequence 534 AA;
 QY Query Match 1.2%; Score 8; DB 2; Length 534;
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 605 DRADSRSS 612
 78 DRADSRSS 85
 RESULT 49
 AAR04920
 ID AAR04920 standard; protein; 549 AA.
 AC AAR04920;
 XX
 XX 31-OCT-2002 (revised)
 DT 02-OCT-1990 (first entry)
 XX
 DE Immunoprotein PEX46.
 XX
 XX Soluble T4 protein; immunotoxin; Pseudomonas endotoxin A; AIDS; HIV; ARC;
 KM angiosenin; fusion protein; PEX46; PEX46.
 KW
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Region 2..182
 FT /label=AAs 3-183 of T4 protein
 FT 183..549
 FT Region /label=AAs 253-613 of Pseudomonas exotoxin A (PE40)
 FT
 FN WO9004414-A.
 XX
 XX 03-MAY-1990.
 PD
 XX 18-OCT-1988; 88US-00259355.
 PF
 XX 18-OCT-1988; 88US-00259355.
 PR
 XX (BIOJ) BIOGEN INC.
 PA
 PI Meade HM, Lobb RR, Gates LL, Winkler G;
 XX
 DR WPI; 1990-163876/21.
 XX
 PT New immunotoxin contg. soluble T4 protein components and toxin - esp.
 PT Pseudomonas endotoxin A, for treating or controlling AIDS and related
 PT conditions, and new DNA sequences.
 XX
 PS Disclosure; Page ?; -pp; English.
 XX
 CC This fusion immunoprotein was produced by constructing a hybrid DNA
 CC sequence of a soluble T4 protein and a fragment of Pseudomonas exotoxin A
 CC (PE40) in which all binding region of PE was removed and which did not
 CC contain the boundary region. The hybrid DNA can then be inserted into an
 CC expression vector and used to produce recombinant fusion protein which is
 CC useful for preventing or treating AIDS, ARC, and HIV infections. The T4
 CC protein is an HIV receptor which binds to the virus or to infected cells

CC carrying the gp120/160 marker antigen, so provides v. specific targeting
 CC with minimal damage to non-target cells. Unmodified chain of PEX45 has a
 CC Mol. Wt. of 59,658. (Updated on 31-OCT-2002 to add missing OS field.)

XX SQ Sequence 549 AA;

Query Match 1.2%; Score 8; DB 2; Length 549;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 DRADSRRS 612
 |||||

DB 54 DRADSRRS 61

RESULT 50

ID AAR04923 standard; protein; 557 AA.

XX AAR04923;

DT 31-OCT-2002 (revised)

DT 02-OCT-1990 (first entry)

DE Immunoprotein TANG11.

KM Soluble T4 protein; immunotoxin; Pseudomonas endotoxin A; AIDS; HIV; ARC;

KW angiogenin; fusion protein; TANG11.

OS Synthetic.

FH Key Location/Qualifiers

FT Region 1..182 /label= AAs 3-183 of T4 protein

FT Region 184..300 /label= AAs 3-183 of T4 protein

FT Region 301..340 /label= translocation domain of PE

FT Region 341..429 /label= binding domain of PE

FT Region 430..559 /label= linker from ADP ribosylation domain of PE

FT Region 433..559 /label= mature angiogenin

PN WO9004414-A.

PD 03-MAY-1990.

PF 18-OCT-1988; 88US-00259355.

PR 18-OCT-1988; 88US-00259355.

PA (BIOJ) BIOGEN INC.

PI Meade HM, Lobb RR, Gates LL, Winkler G;

DR WPI; 1990-163876/21.

PT New immunotoxin contg. soluble T4 protein components and toxin - esp.
 PT Pseudomonas endotoxin A, for treating or controlling AIDS and related
 PT conditions, and new DNA sequences.

PS Disclosure; Page 7; -pp; English.

CC This fusion immunoprotein was produced by constructing a hybrid DNA
 CC sequence of a soluble T4 protein, a fragment of Pseudomonas exotoxin A
 CC including the translocation domain and part of the ADP ribosylation
 CC domain, and the gene for mature angiogenin. The hybrid DNA can then be
 CC inserted into an expression vector and used to produce recombinant fusion
 CC protein which is useful for preventing or treating AIDS, ARC, and HIV
 CC infections. The T4 protein is an HIV receptor which binds to the virus or
 CC to infected cells carrying the gp120/160 marker Ag, so provides v.
 CC specific targeting with minimal damage to non-target cells. (Updated on
 CC 31-OCT-2002 to add missing OS field.)

XX SQ Sequence 557 AA;

Query Match 1.2%; Score 8; DB 2; Length 557;
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 DRADSRRS 612
 |||||

DB 54 DRADSRRS 61

Search completed: June 21, 2004, 13:26:55
 Job time : 61 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: June 21, 2004, 13:26:59 ; Search time 47 Seconds
(without alignments)
3994.438 Million cell updates/sec

Title: US-10-029-345A-109

Perfect score: 665

Sequence: 1 MAHEMIGTQIVTERLVALL.....LGNVSSQSSFSGSMETIEVS 665

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 1163542 seqs, 28213646 residues

Word size : 0

Total number of hits satisfying chosen parameters: 174829

Minimum DB seq length: 473

Maximum DB seq length: 200000000

Post-processing: Listing first 75 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubppaa/US09C_PUBCOMB.pep:*
4: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep:*
5: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubppaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubppaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubppaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubppaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubppaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubppaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	472	71.0	517	9	US-09-964-277-21
2	472	71.0	665	9	US-09-816-494-2
3	472	71.0	665	9	US-09-964-277-2
4	472	71.0	665	12	US-10-072-012-680
5	472	71.0	665	12	US-10-072-012-681
6	472	71.0	665	12	US-10-168-506-14
7	472	71.0	665	12	US-10-343-357-7
8	472	71.0	665	15	US-10-094-749-2312
9	472	71.0	665	15	US-10-377-072-26
10	472	71.0	665	16	US-10-257-026-2
11	472	71.0	665	16	US-10-648-593-240
12	472	71.0	665	16	US-10-648-593-247
13	472	71.0	690	12	US-10-072-012-679
14	472	71.0	690	12	US-10-072-012-703
15	472	71.0	690	12	US-10-425-114-54204

16	394	59.2	680	12	US-10-072-012-256	Sequence 256, App
17	304	45.7	662	12	US-10-072-012-258	Sequence 258, App
18	258	38.8	672	12	US-10-296-115-1259	Sequence 1259, Ap
19	67	10.1	660	12	US-10-072-012-682	Sequence 682, App
20	64	9.6	677	12	US-10-072-012-683	Sequence 683, App
21	20	3.0	501	12	US-10-072-012-702	Sequence 702, App
22	20	3.0	625	12	US-10-072-012-699	Sequence 699, App
23	20	3.0	663	12	US-10-072-012-700	Sequence 700, App
24	12	1.8	616	12	US-10-072-012-266	Sequence 266, App
25	8	1.2	530	8	US-08-485-163-5	Sequence 5, Appl1
26	8	1.2	530	8	US-09-766-995-4	Sequence 4, Appl1
27	8	1.2	532	10	US-09-939-537-6	Sequence 6, Appl1
28	8	1.2	532	11	US-09-243-008-6	Sequence 6, Appl1
29	8	1.2	544	14	US-10-156-761-11387	Sequence 1187, A
30	8	1.2	575	10	US-09-939-537-4	Sequence 4, Appl1
31	8	1.2	575	11	US-09-243-008-4	Sequence 4, Appl1
32	8	1.2	590	9	US-09-934-060A-13	Sequence 13, Appl
33	8	1.2	720	9	US-09-934-060A-2	Sequence 2, Appl1
34	8	1.2	720	9	US-09-934-060A-4	Sequence 4, Appl1
35	8	1.2	774	12	US-10-424-599-25796	Sequence 25796,
36	8	1.2	788	14	US-10-073-118-26	Sequence 26, Appl
37	7	1.1	475	12	US-10-282-122A-47101	Sequence 47101, A
38	7	1.1	478	12	US-10-424-599-245600	Sequence 245600,
39	7	1.1	479	12	US-10-424-599-209874	Sequence 209874,
40	7	1.1	498	14	US-10-032-585-7589	Sequence 7589, Ap
41	7	1.1	503	12	US-10-425-114-62654	Sequence 62654, A
42	7	1.1	512	16	US-10-355-238-13	Sequence 13, Appl
43	7	1.1	512	16	US-10-355-238-23	Sequence 23, Appl
44	7	1.1	515	16	US-10-355-238-14	Sequence 14, Appl
45	7	1.1	515	16	US-10-355-238-25	Sequence 25, Appl
46	7	1.1	516	14	US-10-156-761-8798	Sequence 8798, Ap
47	7	1.1	523	12	US-10-072-012-732	Sequence 732, App
48	7	1.1	525	12	US-10-282-122A-61404	Sequence 61404, A
49	7	1.1	526	14	US-10-361-460-73	Sequence 103, Appl
50	7	1.1	532	14	US-10-289-757-106	Sequence 106, Appl
51	7	1.1	537	12	US-10-424-599-16496	Sequence 16496,
52	7	1.1	539	9	US-09-771-208-4	Sequence 4, Appl1
53	7	1.1	539	12	US-10-424-599-26414	Sequence 26414,
54	7	1.1	542	14	US-10-289-757-107	Sequence 107, App
55	7	1.1	543	14	US-10-289-757-183	Sequence 183, App
56	7	1.1	544	12	US-10-282-122A-48564	Sequence 48564, A
57	7	1.1	547	12	US-10-425-114-50980	Sequence 50980, A
58	7	1.1	548	12	US-10-282-122A-65704	Sequence 65704, A
59	7	1.1	550	12	US-10-425-114-47928	Sequence 47928, A
60	7	1.1	555	9	US-09-764-868-878	Sequence 878, App
61	7	1.1	559	15	US-10-369-493-19468	Sequence 19468, A
62	7	1.1	560	10	US-09-863-776-69	Sequence 69, Appl
63	7	1.1	561	9	US-09-764-864-1539	Sequence 1539, Ap
64	7	1.1	563	12	US-10-425-114-59259	Sequence 59259, A
65	7	1.1	564	9	US-09-999-248-14	Sequence 14, Appl
66	7	1.1	564	12	US-09-878-722-245	Sequence 245, App
67	7	1.1	564	12	US-09-904-456-245	Sequence 245, App
68	7	1.1	564	14	US-10-157-031-176	Sequence 176, App
69	7	1.1	564	14	US-10-318-906A-2	Sequence 2, Appl1
70	7	1.1	564	14	US-10-319-236A-2	Sequence 2, Appl1
71	7	1.1	564	16	US-10-618-839-2	Sequence 2, Appl1
72	7	1.1	569	12	US-10-425-114-39496	Sequence 39496, A
73	7	1.1	580	14	US-10-156-761-12566	Sequence 12566, A
74	7	1.1	581	12	US-10-425-114-56980	Sequence 56980, A
75	7	1.1	584	14	US-10-029-386-32446	Sequence 32446, A

ALIGNMENTS

RESULT 1
US-09-964-277-21
; Sequence 21, Application US/09964277
; Patent No. US20020137170A1
; GENERAL INFORMATION:
; APPLICANT: Lucche, Ralf M.
; APPLICANT: Wei, Bo
; TITLE OF INVENTION: DSP-16 DUAL-SPECIFICITY PHOSPHATASE

FILE REFERENCE: 200125.434
CURRENT APPLICATION NUMBER: US/09/964.277
CURRENT FILING DATE: 2001-09-25
NUMBER OF SEQ ID NOS: 22
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 21
LENGTH: 517
TYPE: PRT
ORGANISM: Homo sapiens
US-09-964-277-21

Query Match 71.0%; Score 472; DB 9; Length 517;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 194 TCPRDPIPESHFLRVVNDSPCEKILPWLDKSVDFIEKAKASNGCVLVHCLAGISRAT 253
DB 46 TCPRDPIPESHFLRVVNDSPCEKILPWLDKSVDFIEKAKASNGCVLVHCLAGISRAT 105
QY 254 IAIAYIMKRMDSLDEAYRFVKEKRPITSPNPNFLGQLDYEKKIKNOTGASGPKSKL 313
DB 106 IAIAYIMKRMDSLDEAYRFVKEKRPITSPNPNFLGQLDYEKKIKNOTGASGPKSKL 165
QY 314 LHLEKNEPVPAVSEGGQKSETPLSPPCADSATSEAAQRPVHPASVPSVPQSLLED 373
DB 166 LHLEKNEPVPAVSEGGQKSETPLSPPCADSATSEAAQRPVHPASVPSVPQSLLED 225
QY 374 SPLVQALSGHLASADRLSDSNKLRSPSLDIKSVSYASMAASLHGFSSEDALEYKPS 433
DB 226 SPLVQALSGHLASADRLSDSNKLRSPSLDIKSVSYASMAASLHGFSSEDALEYKPS 285
QY 434 TTLDTGNKLCQFSPVQELSEOTPEPSPDKEASIPKKLQTPARPSDSQSKLHSVRTSSG 493
DB 286 TTLDTGNKLCQFSPVQELSEOTPEPSPDKEASIPKKLQTPARPSDSQSKLHSVRTSSG 345
QY 494 TAORSLSPLRSGSVEDNYHTSFLGLSTSQOHLTKSAGLGLKGMHSDIILAPQSTPSL 553
DB 346 TAORSLSPLRSGSVEDNYHTSFLGLSTSQOHLTKSAGLGLKGMHSDIILAPQSTPSL 405
QY 554 TSMWFATESHSHFYASAIYGSASYSAYSCQLPTCGDOYVSVRRQKPSDRADRSRW 613
DB 406 TSMWFATESHSHFYASAIYGSASYSAYSCQLPTCGDOYVSVRRQKPSDRADRSRW 465
QY 614 HEESPEKQFKRRSCQMEFGESIMSENRSREELGVGQSFSFGSMELIEVS 665
DB 466 HEESPEKQFKRRSCQMEFGESIMSENRSREELGVGQSFSFGSMELIEVS 517
```

RESULT 2

US-09-816-494-2
Sequence 2, Application US/09816494
Patent No. US20020034807A1
GENERAL INFORMATION:
APPLICANT: Meyers, Rachel A.
TITLE OF INVENTION: 38692 AND 21117, NOVEL DUAL SPECIFICITY
FILE REFERENCE: 10448-030002
CURRENT APPLICATION NUMBER: US/09/816.494
CURRENT FILING DATE: 2001-03-23
PRIOR APPLICATION NUMBER: US 60/191,858
PRIOR FILING DATE: 2000-03-24
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 665
TYPE: PRT
ORGANISM: Homo sapiens
US-09-816-494-2

Query Match 71.0%; Score 472; DB 9; Length 665;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 194 TCPRDPIPESHFLRVVNDSPCEKILPWLDKSVDFIEKAKASNGCVLVHCLAGISRAT 253
DB 194 TCPRDPIPESHFLRVVNDSPCEKILPWLDKSVDFIEKAKASNGCVLVHCLAGISRAT 253
QY 254 IAIAYIMKRMDSLDEAYRFVKEKRPITSPNPNFLGQLDYEKKIKNOTGASGPKSKL 313
DB 254 IAIAYIMKRMDSLDEAYRFVKEKRPITSPNPNFLGQLDYEKKIKNOTGASGPKSKL 313
QY 314 LHLEKNEPVPAVSEGGQKSETPLSPPCADSATSEAAQRPVHPASVPSVPQSLLED 373
DB 314 LHLEKNEPVPAVSEGGQKSETPLSPPCADSATSEAAQRPVHPASVPSVPQSLLED 373
QY 374 SPLVQALSGHLASADRLSDSNKLRSPSLDIKSVSYASMAASLHGFSSEDALEYKPS 433
DB 374 SPLVQALSGHLASADRLSDSNKLRSPSLDIKSVSYASMAASLHGFSSEDALEYKPS 433
QY 434 TTLDTGNKLCQFSPVQELSEOTPEPSPDKEASIPKKLQTPARPSDSQSKLHSVRTSSG 493
DB 434 TTLDTGNKLCQFSPVQELSEOTPEPSPDKEASIPKKLQTPARPSDSQSKLHSVRTSSG 493
QY 494 TAORSLSPLRSGSVEDNYHTSFLGLSTSQOHLTKSAGLGLKGMHSDIILAPQSTPSL 553
DB 494 TAORSLSPLRSGSVEDNYHTSFLGLSTSQOHLTKSAGLGLKGMHSDIILAPQSTPSL 553
QY 554 TSMWFATESHSHFYASAIYGSASYSAYSCQLPTCGDOYVSVRRQKPSDRADRSRW 613
DB 554 TSMWFATESHSHFYASAIYGSASYSAYSCQLPTCGDOYVSVRRQKPSDRADRSRW 613
QY 614 HEESPEKQFKRRSCQMEFGESIMSENRSREELGVGQSFSFGSMELIEVS 665
DB 614 HEESPEKQFKRRSCQMEFGESIMSENRSREELGVGQSFSFGSMELIEVS 665
```

RESULT 3

US-09-964-277-2
Sequence 2, Application US/09964277
Patent No. US20020137170A1
GENERAL INFORMATION:
APPLICANT: Luche, Ralf M.
TITLE OF INVENTION: DSP-16 DUAL-SPECIFICITY PHOSPHATASE
FILE REFERENCE: 200125.434
CURRENT APPLICATION NUMBER: US/09/964.277
CURRENT FILING DATE: 2001-09-25
NUMBER OF SEQ ID NOS: 22
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 665
TYPE: PRT
ORGANISM: Homo sapiens
US-09-964-277-2

Query Match 71.0%; Score 472; DB 9; Length 665;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 194 TCPRDPIPESHFLRVVNDSPCEKILPWLDKSVDFIEKAKASNGCVLVHCLAGISRAT 253
DB 194 TCPRDPIPESHFLRVVNDSPCEKILPWLDKSVDFIEKAKASNGCVLVHCLAGISRAT 253
QY 254 IAIAYIMKRMDSLDEAYRFVKEKRPITSPNPNFLGQLDYEKKIKNOTGASGPKSKL 313
DB 254 IAIAYIMKRMDSLDEAYRFVKEKRPITSPNPNFLGQLDYEKKIKNOTGASGPKSKL 313
QY 314 LHLEKNEPVPAVSEGGQKSETPLSPPCADSATSEAAQRPVHPASVPSVPQSLLED 373
DB 314 LHLEKNEPVPAVSEGGQKSETPLSPPCADSATSEAAQRPVHPASVPSVPQSLLED 373
QY 374 SPLVQALSGHLASADRLSDSNKLRSPSLDIKSVSYASMAASLHGFSSEDALEYKPS 433
DB 374 SPLVQALSGHLASADRLSDSNKLRSPSLDIKSVSYASMAASLHGFSSEDALEYKPS 433
QY 434 TTLDTGNKLCQFSPVQELSEOTPEPSPDKEASIPKKLQTPARPSDSQSKLHSVRTSSG 493
```

```
Db 434 TTLDGNTKLCQFSPVOELSEQTPETSPDKEBASIPKKLQOTARPSDSQSKRLHSVRTSSSG 493
Qy 494 TAQRSLSLPLHRSAGSVEDNYHTSFLFGLSTSOOHLTKSAGLKGWHSIDLAPQISTPBL 553
Db 494 TAQRSLSLPLHRSAGSVEDNYHTSFLFGLSTSOOHLTKSAGLKGWHSIDLAPQISTPBL 553
Qy 554 TSSWYFATSSHPYASAIYGGASAYSAYSCQOLPTCGQOYVSVRROKPSDRADRRSM 613
Db 554 TSSWYFATSSHPYASAIYGGASAYSAYSCQOLPTCGQOYVSVRROKPSDRADRRSM 613
Qy 614 HESPEKQFKRRSCOMERGESIMSENRSREELGKVGSQSSPFGSMELIIVS 665
Db 614 HESPEKQFKRRSCOMERGESIMSENRSREELGKVGSQSSPFGSMELIIVS 665

RESULT 4
US-10-072-012-680
; Sequence 680, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zethusen, Bryan
; APPLICANT: Paturajan, Meera
; APPLICANT: Shinkete, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangoli, Baha
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie E.
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Colman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grose, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See file Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 680
; LENGTH: 665
```

```
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-012-680
Query Match 71.0%; Score 472; DB 12; Length.665;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0,

Qy 194 TCPKPDFIPESHFLRPVNVDSFCEKILPWLDSVDFIEKAKSNGCVLVHCLAGISRST 253
Db 194 TCPKPDFIPESHFLRPVNVDSFCEKILPWLDSVDFIEKAKSNGCVLVHCLAGISRST 253
Qy 254 IAIAYIMKMDNSLDEAYRFEVKEKPTTISPENFNLGQLDYEEKIKNOTGASGPKSLX 313
Db 254 IAIAYIMKMDNSLDEAYRFEVKEKPTTISPENFNLGQLDYEEKIKNOTGASGPKSLX 313
Qy 314 LHLERKNEPVPVAVSEGGKSETPLSPPCADSATSEAAAGRPVHPASVPSPVQPSLLBD 373
Db 314 LHLERKNEPVPVAVSEGGKSETPLSPPCADSATSEAAAGRPVHPASVPSPVQPSLLBD 373
Qy 374 SPLVQALSGIHLSDRLSDSNLTKSFSLDITSVYSASMAASLHGFSSSEDALEYKPS 433
Db 374 SPLVQALSGIHLSDRLSDSNLTKSFSLDITSVYSASMAASLHGFSSSEDALEYKPS 433
Qy 434 TTLDGNTKLCQFSPVOELSEQTPETSPDKEBASIPKKLQOTARPSDSQSKRLHSVRTSSSG 493
Db 434 TTLDGNTKLCQFSPVOELSEQTPETSPDKEBASIPKKLQOTARPSDSQSKRLHSVRTSSSG 493
Qy 494 TAQRSLSLPLHRSAGSVEDNYHTSFLFGLSTSOOHLTKSAGLKGWHSIDLAPQISTPBL 553
Db 494 TAQRSLSLPLHRSAGSVEDNYHTSFLFGLSTSOOHLTKSAGLKGWHSIDLAPQISTPBL 553
Qy 554 TSSWYFATSSHPYASAIYGGASAYSAYSCQOLPTCGQOYVSVRROKPSDRADRRSM 613
Db 554 TSSWYFATSSHPYASAIYGGASAYSAYSCQOLPTCGQOYVSVRROKPSDRADRRSM 613
Qy 614 HESPEKQFKRRSCOMERGESIMSENRSREELGKVGSQSSPFGSMELIIVS 665
Db 614 HESPEKQFKRRSCOMERGESIMSENRSREELGKVGSQSSPFGSMELIIVS 665

RESULT 5
US-10-072-012-681
; Sequence 681, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zethusen, Bryan
; APPLICANT: Paturajan, Meera
; APPLICANT: Shinkete, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangoli, Baha
; APPLICANT: Padigar, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie E.
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Colman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grose, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; CURRENT FILING DATE: 2002-01-31
```

```

; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 681
; LENGTH: 665
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-012-681

Query Match      71.0%; Score 472; DB 12; Length 665;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      194  TCPKPFIPESHFLRPVNDSECEKILPMIDKSVDFIEKAKASNGCVVHCLAGISRAT 253
DB      194  TCPKPFIPESHFLRPVNDSECEKILPMIDKSVDFIEKAKASNGCVVHCLAGISRAT 253
QY      254  IAIAYIMKMDMSLDEAYRFVKEKRPPTISPNFNLGQLLDYEKKIKNOTGASGPKSKLKL 313
DB      254  IAIAYIMKMDMSLDEAYRFVKEKRPPTISPNFNLGQLLDYEKKIKNOTGASGPKSKLKL 313
QY      314  LHEKNEBPVAVSEGGQKSETPSPCADSATSEAGORPVHPASVPSVPSVQPSLLED 373
DB      314  LHEKNEBPVAVSEGGQKSETPSPCADSATSEAGORPVHPASVPSVPSVQPSLLED 373
QY      374  SPLVQALSGHLISADRLSDSNKLKRSFSLDIKSVSYASMAASLHGFSSEDALEYKPS 433
DB      374  SPLVQALSGHLISADRLSDSNKLKRSFSLDIKSVSYASMAASLHGFSSEDALEYKPS 433
QY      434  TTLDTGNKLCQFSPVOELSEQTPETSPDKEASIPKQLQTPARPSDSQSKRLHSVRTSSSG 493
DB      434  TTLDTGNKLCQFSPVOELSEQTPETSPDKEASIPKQLQTPARPSDSQSKRLHSVRTSSSG 493
QY      494  TAORSLSPLHRSGSVEDNYHTSFLFGLSTSOQHLITKSAGLGLKGMHSDIILAPQSTPSL 553
DB      494  TAORSLSPLHRSGSVEDNYHTSFLFGLSTSOQHLITKSAGLGLKGMHSDIILAPQSTPSL 553
QY      554  TSSWYFATESSHFYASAIYGGASAYAGSCQLPTCGDQVYSVRRQKPSDRADSRBSW 613
DB      554  TSSWYFATESSHFYASAIYGGASAYAGSCQLPTCGDQVYSVRRQKPSDRADSRBSW 613
QY      614  HESPEPEKQFKRRSCQMEFGESIMSENRREELGKVGSSQSFSGSMETIEVS 665
DB      614  HESPEPEKQFKRRSCQMEFGESIMSENRREELGKVGSSQSFSGSMETIEVS 665

RESULT 6
US-10-168-506-14
; Sequence 14, Application US/10168506
; Publication No. US20040053229A1
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY D.
; APPLICANT: MARTINEZ, RICARDO

```

```

; APPLICANT: WHYTE, DAVID
; APPLICANT: MANNING, GERARD
; APPLICANT: SUDARSANAM, SUCHA
; APPLICANT: HILL, RON
; APPLICANT: FLANAGAN, PETER
; TITLE OF INVENTION: MAMMALIAN PROTEIN PHOSPHATASES
; FILE REFERENCE: 036602/1351
; CURRENT APPLICATION NUMBER: US/10/168,506
; PRIOR FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: PCT/US00/34736
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 665
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-168-506-14

Query Match      71.0%; Score 472; DB 12; Length 665;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      194  TCPKPFIPESHFLRPVNDSECEKILPMIDKSVDFIEKAKASNGCVVHCLAGISRAT 253
DB      194  TCPKPFIPESHFLRPVNDSECEKILPMIDKSVDFIEKAKASNGCVVHCLAGISRAT 253
QY      254  IAIAYIMKMDMSLDEAYRFVKEKRPPTISPNFNLGQLLDYEKKIKNOTGASGPKSKLKL 313
DB      254  IAIAYIMKMDMSLDEAYRFVKEKRPPTISPNFNLGQLLDYEKKIKNOTGASGPKSKLKL 313
QY      314  LHEKNEBPVAVSEGGQKSETPSPCADSATSEAGORPVHPASVPSVPSVQPSLLED 373
DB      314  LHEKNEBPVAVSEGGQKSETPSPCADSATSEAGORPVHPASVPSVPSVQPSLLED 373
QY      374  SPLVQALSGHLISADRLSDSNKLKRSFSLDIKSVSYASMAASLHGFSSEDALEYKPS 433
DB      374  SPLVQALSGHLISADRLSDSNKLKRSFSLDIKSVSYASMAASLHGFSSEDALEYKPS 433
QY      434  TTLDTGNKLCQFSPVOELSEQTPETSPDKEASIPKQLQTPARPSDSQSKRLHSVRTSSSG 493
DB      434  TTLDTGNKLCQFSPVOELSEQTPETSPDKEASIPKQLQTPARPSDSQSKRLHSVRTSSSG 493
QY      494  TAORSLSPLHRSGSVEDNYHTSFLFGLSTSOQHLITKSAGLGLKGMHSDIILAPQSTPSL 553
DB      494  TAORSLSPLHRSGSVEDNYHTSFLFGLSTSOQHLITKSAGLGLKGMHSDIILAPQSTPSL 553
QY      554  TSSWYFATESSHFYASAIYGGASAYAGSCQLPTCGDQVYSVRRQKPSDRADSRBSW 613
DB      554  TSSWYFATESSHFYASAIYGGASAYAGSCQLPTCGDQVYSVRRQKPSDRADSRBSW 613
QY      614  HESPEPEKQFKRRSCQMEFGESIMSENRREELGKVGSSQSFSGSMETIEVS 665
DB      614  HESPEPEKQFKRRSCQMEFGESIMSENRREELGKVGSSQSFSGSMETIEVS 665

RESULT 7
US-10-343-357-7
; Sequence 7, Application US/10343357
; Publication No. US20040058341A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.; TANG, Y.Tom
; APPLICANT: ELLIOTT, Vicki S.; RAMKUMAR, Jayalamsi
; APPLICANT: YAO, Monique G.; BURFORD, Neil
; APPLICANT: WANG, Yumei E.; STEWART, Elizabeth A.
; APPLICANT: GANDHI, Ameena R.; ARVIZU, Chandra S.
; APPLICANT: LEE, Ernestine A.; HAPALIA, April J.A.
; APPLICANT: LU, Dying Aina M.; TRIBOUDEY, Catherine M.
; APPLICANT: GRIFFIN, Jennifer A.; BAUGHN, Mariah R.
; APPLICANT: NGUYEN, Henry; WARREN, Bridget A.
; APPLICANT: NGUYEN, Daniel B.; CHAMLA, Narinder K.
; APPLICANT: KEARNEY, Liam
; TITLE OF INVENTION: PROTEIN PHOSPHATASES

```

```

; APPLICANT: MAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: MAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHICO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTUYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH CDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2312
; LENGTH: 665
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2312

Query Match          71.0%; Score 472; DB 15; Length 665;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      194  TCGPRDPRIPESHFLRVPVNDSPFCCKILPMLDKSVDFITKAKASNGCVLVHCLAGISRSAT 253
DB      194  TCGPRDPRIPESHFLRVPVNDSPFCCKILPMLDKSVDFITKAKASNGCVLVHCLAGISRSAT 253
QY      254  IAIYIYIMKRDMSLDEAIFYEKRPRTISPNFNLGOLLIDYEKKIKNOTGASGPKSKLKL 313
DB      254  IAIYIYIMKRDMSLDEAIFYEKRPRTISPNFNLGOLLIDYEKKIKNOTGASGPKSKLKL 313
QY      314  LHLKEPNEPVAVSEGGQKSETPLSPPCADSATSEBAQGRPVHNPASVPSVPSLLED 373
DB      314  LHLKEPNEPVAVSEGGQKSETPLSPPCADSATSEBAQGRPVHNPASVPSVPSLLED 373
QY      374  SPLVQALISGLHSLADRLSDSNKLRKRSFSLDKSVYSASMAASLHGFSSEDALEYTPRS 433
DB      374  SPLVQALISGLHSLADRLSDSNKLRKRSFSLDKSVYSASMAASLHGFSSEDALEYTPRS 433
QY      434  TTLGGINKLCOFSPVOELSEOTPEPSPDKKEASIPKYLQOTARPSPDSOSKRLHSVYRTSSG 493
DB      434  TTLGGINKLCOFSPVOELSEOTPEPSPDKKEASIPKYLQOTARPSPDSOSKRLHSVYRTSSG 493
QY      494  TAQRSLSPPLHRSGSVEDNHYTSELFGLSTSQOHLTKSAGLKGWHSDDLAPOTSTPSL 553
DB      494  TAQRSLSPPLHRSGSVEDNHYTSELFGLSTSQOHLTKSAGLKGWHSDDLAPOTSTPSL 553
QY      554  TSSWYFATESHFTYSASAIYGGASYSAYSCSGLPTCGDQVYSVRRROKPSDRADSRSM 613
DB      554  TSSWYFATESHFTYSASAIYGGASYSAYSCSGLPTCGDQVYSVRRROKPSDRADSRSM 613
QY      614  HEEBPFEKQFRRRSQMEFGESINSENRRELDKVGSGSFFSGSMETIEVS 665
DB      614  HEEBPFEKQFRRRSQMEFGESINSENRRELDKVGSGSFFSGSMETIEVS 665

RESULT 9
US-10-377-072-26
; Sequence 26, Application US/10377072
; Publication No. US20040009501AI
; GENERAL INFORMATION:

```



```
/ APPLICANT: Millennium Pharmaceuticals Inc.
/ APPLICANT: Curtis, Rory A.J.
/ APPLICANT: Logan, Thomas Joseph
/ APPLICANT: Glucksmann, Maria A.
/ APPLICANT: Meyers, Rachel E.
/ APPLICANT: Williams, Mark J.
/ APPLICANT: Rudolph-Owen, Laura A.
/ APPLICANT: Chun, Miyoung
/ APPLICANT: Tsai, Fong-Ying
/ TITLE OF INVENTION: NOVEL 25869, 25934, 26335, 50365, 21117,
/ TITLE OF INVENTION: 38692, 46508, 16816, 16839, 49937, 49931 AND 49933 MOLECULES
/ TITLE OF INVENTION: AND USES THEREFOR
/ FILE REFERENCE: MP103-0180MIM
/ CURRENT APPLICATION NUMBER: US/10/377,072
/ CURRENT FILING DATE: 2003-02-27
/ PRIOR APPLICATION NUMBER: US 09/895,860
/ PRIOR FILING DATE: 2001-06-29
/ PRIOR APPLICATION NUMBER: US 60/215,370
/ PRIOR FILING DATE: 2000-06-29
/ PRIOR APPLICATION NUMBER: US 09/723,806
/ PRIOR FILING DATE: 2000-11-28
/ PRIOR APPLICATION NUMBER: US 60/187,455
/ PRIOR FILING DATE: 2000-03-07
/ PRIOR APPLICATION NUMBER: US 09/843,297
/ PRIOR FILING DATE: 2001-04-25
/ PRIOR APPLICATION NUMBER: US 60/199,801
/ PRIOR FILING DATE: 2000-04-26
/ PRIOR APPLICATION NUMBER: US 09/861,801
/ PRIOR FILING DATE: 2001-05-21
/ PRIOR APPLICATION NUMBER: US 60/205,508
/ PRIOR FILING DATE: 2000-05-19
/ PRIOR APPLICATION NUMBER: US 09/816,494
/ PRIOR FILING DATE: 2001-03-23
/ PRIOR APPLICATION NUMBER: US 09/815,419
/ PRIOR FILING DATE: 2001-03-22
/ Remaining prior Application data removed - See file wrapper or PAM.
/ NUMBER OF SEQ ID NOS: 114
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 26
/ LENGTH: 665
/ TYPE: PRT
/ ORGANISM: Homo Sapiens
/ US-10-377-072-26

Query Match      71.0%; Score 472; DB 15; Length 665;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

194 TCPKPDFIPESHFLRPVNDSEFCCEKILPWLKSDVDFIEKAKANGCVLHCLAGISRAT 253
|||||
194 TCPKPDFIPESHFLRPVNDSEFCCEKILPWLKSDVDFIEKAKANGCVLHCLAGISRAT 253
254 IAIAYIMKMDMSLDEAYRFVKEKRPPTISPNFNFGLQLLDYEEKIKNOTGASGPKSKL 313
|||||
254 IAIAYIMKMDMSLDEAYRFVKEKRPPTISPNFNFGLQLLDYEEKIKNOTGASGPKSKL 313
314 LHLKPNBPVPVAVSEGGCKSETPLSPPCADSATSEAGORPVHPASVPSVPSVPSLLED 373
|||||
314 LHLKPNBPVPVAVSEGGCKSETPLSPPCADSATSEAGORPVHPASVPSVPSVPSLLED 373
314 LHLKPNBPVPVAVSEGGCKSETPLSPPCADSATSEAGORPVHPASVPSVPSVPSLLED 373
374 SPLVQALSGHLASADRLSDSNKLKRSFSLDIKSVSYASMAASLHGFSSESDALEYKPS 433
|||||
374 SPLVQALSGHLASADRLSDSNKLKRSFSLDIKSVSYASMAASLHGFSSESDALEYKPS 433
434 TTLDTGNKLCOFSPVOELSEQTPETSPDKEASIPKQLQTARPSDSQSKRLHSVRTSSG 493
|||||
434 TTLDTGNKLCOFSPVOELSEQTPETSPDKEASIPKQLQTARPSDSQSKRLHSVRTSSG 493
494 TAORSLSPLHRSGSVEDNYHTSFLFGISTSOQHLTKSAGLKGWMSDILAPQTSPTSL 553
|||||
494 TAORSLSPLHRSGSVEDNYHTSFLFGISTSOQHLTKSAGLKGWMSDILAPQTSPTSL 553
554 TSSWTFATESHFPYASAIYGGASYSAYSCSQLPTCGDQVYVRRROKPEDRADSRSM 613
|||||
554 TSSWTFATESHFPYASAIYGGASYSAYSCSQLPTCGDQVYVRRROKPEDRADSRSM 613
```

```
|||||
Db      554 TSSWTFATESHFPYASAIYGGASYSAYSCSQLPTCGDQVYVRRROKPEDRADSRSM 613
|||||
Qy      614 HESPEKQFKRRSCOMERGESIMENSRREBLGAVGSSQSPSGMEIIEVS 665
|||||
Db      614 HESPEKQFKRRSCOMERGESIMENSRREBLGAVGSSQSPSGMEIIEVS 665

RESULT 10
US-10-257-026-2
/ Sequence 2, Application US/10257026
/ Publication No. US20040086859A1
/ GENERAL INFORMATION:
/ APPLICANT: Merck Patent GmbH
/ TITLE OF INVENTION: New dual specificity phosphatase
/ FILE REFERENCE: DUSP10KDM5
/ CURRENT APPLICATION NUMBER: US/10/257,026
/ CURRENT FILING DATE: 2003-11-07
/ NUMBER OF SEQ ID NOS: 3
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 2
/ LENGTH: 665
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ US-10-257-026-2

Query Match      71.0%; Score 472; DB 16; Length 665;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

194 TCPKPDFIPESHFLRPVNDSEFCCEKILPWLKSDVDFIEKAKANGCVLHCLAGISRAT 253
|||||
194 TCPKPDFIPESHFLRPVNDSEFCCEKILPWLKSDVDFIEKAKANGCVLHCLAGISRAT 253
254 IAIAYIMKMDMSLDEAYRFVKEKRPPTISPNFNFGLQLLDYEEKIKNOTGASGPKSKL 313
|||||
254 IAIAYIMKMDMSLDEAYRFVKEKRPPTISPNFNFGLQLLDYEEKIKNOTGASGPKSKL 313
314 LHLKPNBPVPVAVSEGGCKSETPLSPPCADSATSEAGORPVHPASVPSVPSVPSLLED 373
|||||
314 LHLKPNBPVPVAVSEGGCKSETPLSPPCADSATSEAGORPVHPASVPSVPSVPSLLED 373
314 LHLKPNBPVPVAVSEGGCKSETPLSPPCADSATSEAGORPVHPASVPSVPSVPSLLED 373
374 SPLVQALSGHLASADRLSDSNKLKRSFSLDIKSVSYASMAASLHGFSSESDALEYKPS 433
|||||
374 SPLVQALSGHLASADRLSDSNKLKRSFSLDIKSVSYASMAASLHGFSSESDALEYKPS 433
434 TTLDTGNKLCOFSPVOELSEQTPETSPDKEASIPKQLQTARPSDSQSKRLHSVRTSSG 493
|||||
434 TTLDTGNKLCOFSPVOELSEQTPETSPDKEASIPKQLQTARPSDSQSKRLHSVRTSSG 493
494 TAORSLSPLHRSGSVEDNYHTSFLFGISTSOQHLTKSAGLKGWMSDILAPQTSPTSL 553
|||||
494 TAORSLSPLHRSGSVEDNYHTSFLFGISTSOQHLTKSAGLKGWMSDILAPQTSPTSL 553
554 TSSWTFATESHFPYASAIYGGASYSAYSCSQLPTCGDQVYVRRROKPEDRADSRSM 613
|||||
554 TSSWTFATESHFPYASAIYGGASYSAYSCSQLPTCGDQVYVRRROKPEDRADSRSM 613
614 HESPEKQFKRRSCOMERGESIMENSRREBLGAVGSSQSPSGMEIIEVS 665
|||||
614 HESPEKQFKRRSCOMERGESIMENSRREBLGAVGSSQSPSGMEIIEVS 665

RESULT 11
US-10-648-593-240
/ Sequence 240, Application US/10648593
/ Publication No. US20040106132A1
/ GENERAL INFORMATION:
/ APPLICANT: Bristol-Myers Squibb Company
/ TITLE OF INVENTION: IDENTIFICATION OF GENES FOR PREDICTING ACTIVITY OF COMPOUNDS THAT
/ INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE KINASES AND/OR
/ TITLE OF INVENTION: PROTEIN TYROSINE KINASE PATHWAYS IN BREAST CELLS
/ FILE REFERENCE: D0273 NP
```

CURRENT APPLICATION NUMBER: US/10/648,593
CURRENT FILING DATE: 2003-08-26
PRIOR APPLICATION NUMBER: 60/406,385
PRIOR FILING DATE: 2002-08-27
NUMBER OF SEQ ID NOS: 557
SOFTWARE: PatentIn version 3.2
SEQ ID NO 240
LENGTH: 665
TYPE: PRT
ORGANISM: Homo sapiens
US-10-648-593-240

Query Match
Best Local Similarity 100.0%; Pred. No. 0;
Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 TCCKPDPFIPESHFLRVVNDVDFCEKILPWLKSVDFIEKAKASNGCVLVHCLAGISRAT 253
DB 194 TCCKPDPFIPESHFLRVVNDVDFCEKILPWLKSVDFIEKAKASNGCVLVHCLAGISRAT 253
QY 254 IAIAYIMKMDMSLDEAYRFVVEKEKPTTISPNNFQLQQLDYEEKIKNOTGASGPKSKLXL 313
DB 254 IAIAYIMKMDMSLDEAYRFVVEKEKPTTISPNNFQLQQLDYEEKIKNOTGASGPKSKLXL 313
QY 314 LHLEKNEPVPAVSEGGQKSETPSPCADSATSEAAAGORPVHPASVPSPVQPSLLBD 373
DB 314 LHLEKNEPVPAVSEGGQKSETPSPCADSATSEAAAGORPVHPASVPSPVQPSLLBD 373
QY 374 SPLVQALSGHLASADRLSDSNLKRFSLDIKSVYSASMAASLHGFSSEDALEYKKS 433
DB 374 SPLVQALSGHLASADRLSDSNLKRFSLDIKSVYSASMAASLHGFSSEDALEYKKS 433
QY 434 TLLDGTNKLCOFSPVOELSEOTPEPSPDKEASIPKKLQTPARPSDSQKRLHSVRTSSSG 493
DB 434 TLLDGTNKLCOFSPVOELSEOTPEPSPDKEASIPKKLQTPARPSDSQKRLHSVRTSSSG 493
QY 494 TAORSLSPLHSGSVEDNYHTSFLGLSTSOOHLTKSAGLKGWHSIDLAPOTSTPSL 553
DB 494 TAORSLSPLHSGSVEDNYHTSFLGLSTSOOHLTKSAGLKGWHSIDLAPOTSTPSL 553
QY 554 TSSWYPATESSHFYASAIYGSASYSAYSCSQLPTCGQOVSVRROKPSDRADSRSM 613
DB 554 TSSWYPATESSHFYASAIYGSASYSAYSCSQLPTCGQOVSVRROKPSDRADSRSM 613
QY 614 HESPEKQFKRRSCOMERGESIMSENREELGKVGSGSSFGSMETIEVS 665
DB 614 HESPEKQFKRRSCOMERGESIMSENREELGKVGSGSSFGSMETIEVS 665

RESULT 12

US-10-648-593-247
Sequence 247, Application US/10648593
Publication No. US20040106132A1
GENERAL INFORMATION:
APPLICANT: Bristol-Myers Squibb Company
TITLE OF INVENTION: IDENTIFICATION OF GENES FOR PREDICTING ACTIVITY OF COMPOUNDS THAT INTERACT WITH AND/OR MODULATE PROTEIN TYROSINE KINASES AND/OR
TITLE OF INVENTION: PROTEIN TYROSINE KINASE PATHWAYS IN BREAST CELLS
FILE REFERENCE: D0273 NP
CURRENT APPLICATION NUMBER: US/10/648,593
CURRENT FILING DATE: 2003-08-26
PRIOR APPLICATION NUMBER: 60/406,385
PRIOR FILING DATE: 2002-08-27
NUMBER OF SEQ ID NOS: 557
SOFTWARE: PatentIn version 3.2
SEQ ID NO 247
LENGTH: 665
TYPE: PRT
ORGANISM: Homo sapiens
US-10-648-593-247

Query Match
Best Local Similarity 100.0%; Pred. No. 0;
Matches 472; DB 16; Length 665;

Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 TCCKPDPFIPESHFLRVVNDVDFCEKILPWLKSVDFIEKAKASNGCVLVHCLAGISRAT 253
DB 194 TCCKPDPFIPESHFLRVVNDVDFCEKILPWLKSVDFIEKAKASNGCVLVHCLAGISRAT 253
QY 254 IAIAYIMKMDMSLDEAYRFVVEKEKPTTISPNNFQLQQLDYEEKIKNOTGASGPKSKLXL 313
DB 254 IAIAYIMKMDMSLDEAYRFVVEKEKPTTISPNNFQLQQLDYEEKIKNOTGASGPKSKLXL 313
QY 314 LHLEKNEPVPAVSEGGQKSETPSPCADSATSEAAAGORPVHPASVPSPVQPSLLBD 373
DB 314 LHLEKNEPVPAVSEGGQKSETPSPCADSATSEAAAGORPVHPASVPSPVQPSLLBD 373
QY 374 SPLVQALSGHLASADRLSDSNLKRFSLDIKSVYSASMAASLHGFSSEDALEYKKS 433
DB 374 SPLVQALSGHLASADRLSDSNLKRFSLDIKSVYSASMAASLHGFSSEDALEYKKS 433
QY 434 TLLDGTNKLCOFSPVOELSEOTPEPSPDKEASIPKKLQTPARPSDSQKRLHSVRTSSSG 493
DB 434 TLLDGTNKLCOFSPVOELSEOTPEPSPDKEASIPKKLQTPARPSDSQKRLHSVRTSSSG 493
QY 494 TAORSLSPLHSGSVEDNYHTSFLGLSTSOOHLTKSAGLKGWHSIDLAPOTSTPSL 553
DB 494 TAORSLSPLHSGSVEDNYHTSFLGLSTSOOHLTKSAGLKGWHSIDLAPOTSTPSL 553
QY 554 TSSWYPATESSHFYASAIYGSASYSAYSCSQLPTCGQOVSVRROKPSDRADSRSM 613
DB 554 TSSWYPATESSHFYASAIYGSASYSAYSCSQLPTCGQOVSVRROKPSDRADSRSM 613
QY 614 HESPEKQFKRRSCOMERGESIMSENREELGKVGSGSSFGSMETIEVS 665
DB 614 HESPEKQFKRRSCOMERGESIMSENREELGKVGSGSSFGSMETIEVS 665

RESULT 13

US-10-072-012-679
Sequence 679, Application US/10072012
Publication No. US20040033493A1
GENERAL INFORMATION:
APPLICANT: Tchernyev, Valizar
APPLICANT: Spytek, Kimberly
APPLICANT: Zernusen, Bryan
APPLICANT: Paturejan, Meera
APPLICANT: Shimets, Richard
APPLICANT: Li, Li
APPLICANT: Gangoli, Esha
APPLICANT: Padigar, Muralidhara
APPLICANT: Anderson, David W.
APPLICANT: Rastelli, Luca
APPLICANT: Miller, Charles E.
APPLICANT: Gerlach, Valerie
APPLICANT: Taupier Jr, Raymond J.
APPLICANT: Gusev, Vladimir Y.
APPLICANT: Coleman, Steven D.
APPLICANT: Molenc, Adam R.
APPLICANT: Pena, Carol E. A
APPLICANT: Furtak, Katarzyna
APPLICANT: Grosse, William M.
APPLICANT: Alsbrook II, John P.
APPLICANT: Lepley, Denise M.
APPLICANT: Rieger, Daniel K.
APPLICANT: Burgess, Catherine E.
TITLE OF INVENTION: Proteins and Nucleic Acid Encoding Same
FILE REFERENCE: 21402-258
CURRENT APPLICATION NUMBER: US/10/072,012
CURRENT FILING DATE: 2002-01-31
PRIOR APPLICATION NUMBER: 60/265,102
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: 60/265,514
PRIOR FILING DATE: 2001-01-31
PRIOR APPLICATION NUMBER: 60/265,517
PRIOR FILING DATE: 2001-01-31

```
/ PRIOR APPLICATION NUMBER: 60/265,412
/ PRIOR FILING DATE: 2001-01-31
/ PRIOR APPLICATION NUMBER: 60/265,395
/ PRIOR FILING DATE: 2001-01-31
/ PRIOR APPLICATION NUMBER: 60/266,406
/ PRIOR FILING DATE: 2001-02-02
/ PRIOR APPLICATION NUMBER: 60/266,767
/ PRIOR FILING DATE: 2001-02-05
/ PRIOR APPLICATION NUMBER: 60/267,057
/ PRIOR FILING DATE: 2001-02-07
/ PRIOR APPLICATION NUMBER: 60/266,975
/ PRIOR FILING DATE: 2001-02-07
/ PRIOR APPLICATION NUMBER: 60/267,459
/ PRIOR FILING DATE: 2001-02-08
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 1391
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 679
/ LENGTH: 690
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-072-012-679
```

```
Query Match      71.0%; Score 472; DB 12; Length 690;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 194 TCPRPFIPIESHFLRVVNDSPFCEKILPWLDKSVDFIEKAKASNGCVLVHCLAGISRAT 253
DB 219 TCPRPFIPIESHFLRVVNDSPFCEKILPWLDKSVDFIEKAKASNGCVLVHCLAGISRAT 278
QY 254 IAIAYIMKMDMSLDKAYRVVKEKRPITISPNFNLGQLLDYEKKIKNOTGASGPKSKL 313
DB 279 IAIAYIMKMDMSLDKAYRVVKEKRPITISPNFNLGQLLDYEKKIKNOTGASGPKSKL 338
QY 314 LHLEKNEVPVAVSEGGQKSETPPLSPPCADSATSEAAQGPVHPASVPSVQPSLLED 373
DB 339 LHLEKNEVPVAVSEGGQKSETPPLSPPCADSATSEAAQGPVHPASVPSVQPSLLED 398
QY 374 SPLVQALSGHLHSADRLSDNSNKLKRSFSLDIKSVSYASMAASLHGFSSEDALEYKKS 433
DB 399 SPLVQALSGHLHSADRLSDNSNKLKRSFSLDIKSVSYASMAASLHGFSSEDALEYKKS 458
QY 434 TTLDTGNKLCOFSPVQELSEOTPETSPDKEEASIPKKLQOTARPDSOSKRLHVSRTSSG 493
DB 459 TTLDTGNKLCOFSPVQELSEOTPETSPDKEEASIPKKLQOTARPDSOSKRLHVSRTSSG 518
QY 494 TKORSILSLPHRSGSVEDNYHTSFLFGLSTSOQHLTKSAGLGKGMHSDILAPQTSPTSL 553
DB 519 TKORSILSLPHRSGSVEDNYHTSFLFGLSTSOQHLTKSAGLGKGMHSDILAPQTSPTSL 578
QY 554 TSSWYPATESHSHFYASAIYGGASASAYASCOLPTCCGQYVSVRRQKPSRAADSRSM 613
DB 579 TSSWYPATESHSHFYASAIYGGASASAYASCOLPTCCGQYVSVRRQKPSRAADSRSM 638
QY 614 HESPEPEKQFKRRSCOMEPGESIMSENRREELGKYGQSOSFSGMEITIEVS 665
DB 639 HESPEPEKQFKRRSCOMEPGESIMSENRREELGKYGQSOSFSGMEITIEVS 690
```

```
RESULT 14
US-10-072-012-703
/ Sequence 703, Application US/10072012
/ Publication No. US2004003493A1
/ GENERAL INFORMATION:
/ APPLICANT: Tcheinev, Velizar
/ APPLICANT: Spytek, Kimberly
/ APPLICANT: Zernusen, Bryan
/ APPLICANT: Patuturajan, Meera
/ APPLICANT: Shinkets, Richard
/ APPLICANT: Li, Li
/ APPLICANT: Gangolli, Esha
/ APPLICANT: Padigaru, Muralidhara
```

```
/ APPLICANT: Anderson, David W.
/ APPLICANT: Rastelli, Luca
/ APPLICANT: Miller, Charles E.
/ APPLICANT: Gerlach, Valerie
/ APPLICANT: Taudier Jr, Raymond J.
/ APPLICANT: Gusev, Vladimir Y.
/ APPLICANT: Colman, Steven D.
/ APPLICANT: Wolenc, Adam R.
/ APPLICANT: Pena, Carol E. A
/ APPLICANT: Furtak, Katarzyna
/ APPLICANT: Grosse, William M.
/ APPLICANT: Alsobrook II, John P.
/ APPLICANT: Lepley, Denise M.
/ APPLICANT: Rieger, Daniel K.
/ APPLICANT: Burgess, Catherine E.
/ TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
/ FILE REFERENCE: 21402-258
/ CURRENT APPLICATION NUMBER: US/10/072,012
/ PRIOR FILING DATE: 2002-01-31
/ PRIOR APPLICATION NUMBER: 60/265,102
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: 60/265,514
/ PRIOR FILING DATE: 2001-01-31
/ PRIOR APPLICATION NUMBER: 60/265,517
/ PRIOR FILING DATE: 2001-01-31
/ PRIOR APPLICATION NUMBER: 60/265,412
/ PRIOR FILING DATE: 2001-01-31
/ PRIOR APPLICATION NUMBER: 60/265,395
/ PRIOR FILING DATE: 2001-01-31
/ PRIOR APPLICATION NUMBER: 60/266,406
/ PRIOR FILING DATE: 2001-02-02
/ PRIOR APPLICATION NUMBER: 60/266,767
/ PRIOR FILING DATE: 2001-02-05
/ PRIOR APPLICATION NUMBER: 60/267,057
/ PRIOR FILING DATE: 2001-02-07
/ PRIOR APPLICATION NUMBER: 60/266,975
/ PRIOR FILING DATE: 2001-02-07
/ PRIOR APPLICATION NUMBER: 60/267,459
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 1391
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 703
/ LENGTH: 690
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-072-012-703

Query Match      71.0%; Score 472; DB 12; Length 690;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 194 TCPRPFIPIESHFLRVVNDSPFCEKILPWLDKSVDFIEKAKASNGCVLVHCLAGISRAT 253
DB 219 TCPRPFIPIESHFLRVVNDSPFCEKILPWLDKSVDFIEKAKASNGCVLVHCLAGISRAT 278
QY 254 IAIAYIMKMDMSLDKAYRVVKEKRPITISPNFNLGQLLDYEKKIKNOTGASGPKSKL 313
DB 279 IAIAYIMKMDMSLDKAYRVVKEKRPITISPNFNLGQLLDYEKKIKNOTGASGPKSKL 338
QY 314 LHLEKNEVPVAVSEGGQKSETPPLSPPCADSATSEAAQGPVHPASVPSVQPSLLED 373
DB 339 LHLEKNEVPVAVSEGGQKSETPPLSPPCADSATSEAAQGPVHPASVPSVQPSLLED 398
QY 374 SPLVQALSGHLHSADRLSDNSNKLKRSFSLDIKSVSYASMAASLHGFSSEDALEYKKS 433
DB 399 SPLVQALSGHLHSADRLSDNSNKLKRSFSLDIKSVSYASMAASLHGFSSEDALEYKKS 458
QY 434 TTLDTGNKLCOFSPVQELSEOTPETSPDKEEASIPKKLQOTARPDSOSKRLHVSRTSSG 493
DB 459 TTLDTGNKLCOFSPVQELSEOTPETSPDKEEASIPKKLQOTARPDSOSKRLHVSRTSSG 518
QY 494 TKORSILSLPHRSGSVEDNYHTSFLFGLSTSOQHLTKSAGLGKGMHSDILAPQTSPTSL 553
```

Db 519 TAQRSLSPHRSRGSEVDNHTSFLFGLSTSQOHLTKSAGLGLKGMHSDILAQOTSTPSL 578
QY 554 TSSWYFATESSHYYSASAIYGSASYSAYSCSOLPTCGDQVYVRRRQKSPDADSRSSW 613
Db 579 TSSWYFATESSHYYSASAIYGSASYSAYSCSOLPTCGDQVYVRRRQKSPDADSRSSW 638
QY 614 HESPEPKOPKRRSCOMEFGESIMSENREELGKVGSSSFGSGMEIIEVS 665
Db 639 HESPEPKOPKRRSCOMEFGESIMSENREELGKVGSSSFGSGMEIIEVS 690

RESULT 15

US-10-425-114-54204
; Sequence 54204, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jindong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaka, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5311)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 54204
; LENGTH: 690
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB4119-028-H6_FIL pep
US-10-425-114-54204

Query Match 71.0%; Score 472; DB 12; Length 690;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 472; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 194 TCPRPFIPESHFLRVPVNDSPCEKILPMLDKSVDFIEKAKANGCVLHGLAGISRSAT 253
Db 219 TCPRPFIPESHFLRVPVNDSPCEKILPMLDKSVDFIEKAKANGCVLHGLAGISRSAT 278
QY 254 IAIAYIMKMDMSLDEAYRFVKEKRPITISPNFNLGQLDYEKKIKNOTGASGPKSKLKL 313
Db 279 IAIAYIMKMDMSLDEAYRFVKEKRPITISPNFNLGQLDYEKKIKNOTGASGPKSKLKL 338
QY 314 LHEKNEPVPVASEGGQKSETPLSPCCADTSATSEAAQGPVHPASVPSVQPSILED 373
Db 339 LHEKNEPVPVASEGGQKSETPLSPCCADTSATSEAAQGPVHPASVPSVQPSILED 398
QY 374 SPLVQALSGHLGADRLSDNKLKRSFLDIKSVYSASMAALHGFSSSEDLLEYKPS 433
Db 399 SPLVQALSGHLGADRLSDNKLKRSFLDIKSVYSASMAALHGFSSSEDLLEYKPS 458
QY 434 TTLDTGNKLCOFSPVQELSPQTPETSPDKKEASIPKLLQIARPSDSQSKLHSVRTSSSG 493
Db 459 TTLDTGNKLCOFSPVQELSPQTPETSPDKKEASIPKLLQIARPSDSQSKLHSVRTSSSG 518
QY 494 TAQRSLSPHRSRGSEVDNHTSFLFGLSTSQOHLTKSAGLGLKGMHSDILAQOTSTPSL 553
Db 519 TAQRSLSPHRSRGSEVDNHTSFLFGLSTSQOHLTKSAGLGLKGMHSDILAQOTSTPSL 578
QY 554 TSSWYFATESSHYYSASAIYGSASYSAYSCSOLPTCGDQVYVRRRQKSPDADSRSSW 613
Db 579 TSSWYFATESSHYYSASAIYGSASYSAYSCSOLPTCGDQVYVRRRQKSPDADSRSSW 638
QY 614 HESPEPKOPKRRSCOMEFGESIMSENREELGKVGSSSFGSGMEIIEVS 665
Db 639 HESPEPKOPKRRSCOMEFGESIMSENREELGKVGSSSFGSGMEIIEVS 690

RESULT 16

US-10-072-012-256
; Sequence 256, Application US/10072012
; Publication No. US2004003493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zernusen, Bryan
; APPLICANT: Paturajan, Weera
; APPLICANT: Shimkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangoli, Beba
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimyr Y.
; APPLICANT: Colman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grosse, William M.
; APPLICANT: Alsbrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 256
; LENGTH: 680
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-012-256

Query Match 59.2%; Score 394; DB 12; Length 680;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 394; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 272 RFVKEKRPITISPNFNLGQLDYEKKIKNOTGASGPKSKLKLHLEKNEPVPVASEGGQ 331
Db 287 RFVKEKRPITISPNFNLGQLDYEKKIKNOTGASGPKSKLKLHLEKNEPVPVASEGGQ 346
QY 332 KSETPLSPCCADTSATSEAAQGPVHPASVPSVQPSILEDSPVQALSGHLGADRL 391
Db 347 KSETPLSPCCADTSATSEAAQGPVHPASVPSVQPSILEDSPVQALSGHLGADRL 406

```
QY 392 DSNKLRKSFSLDIKVSYSASMAASLHGFSSEDALEYKSTTLTGNTKLCQFSPVOEL 451
|
|
|
Db 407 DSNKLRKSFSLDIKVSYSASMAASLHGFSSEDALEYKSTTLTGNTKLCQFSPVOEL 466
|
|
|
QY 452 SEQTPEKSPKKEASIPKLTQTARPSDSQSKRLHSVTTSSSGTAQBSLSPLRSGSVED 511
|
|
|
Db 467 SEQTPEKSPKKEASIPKLTQTARPSDSQSKRLHSVTTSSSGTAQBSLSPLRSGSVED 526
|
|
|
QY 512 NHTSFLFGLSTSQOHLTKSAGLKGWHSDDLAPQSTPSLTSSWYFATESHFPYSAA 571
|
|
|
Db 527 NHTSFLFGLSTSQOHLTKSAGLKGWHSDDLAPQSTPSLTSSWYFATESHFPYSAA 586
|
|
|
QY 572 IYGSASYSASYSQSLPTCCDQVYSVRRQKPSDRADSRSMHEESPFEKQFKRRSCOME 631
|
|
|
Db 587 IYGSASYSASYSQSLPTCCDQVYSVRRQKPSDRADSRSMHEESPFEKQFKRRSCOME 646
|
|
|
QY 632 FGSIIMSENRSREBELGKVGSSQSFSGSMETIEVS 665
|
|
|
Db 647 FGSIIMSENRSREBELGKVGSSQSFSGSMETIEVS 680
|
|
|
```

RESULT 17

```
US-10-072-012-258
; Sequence 258, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zernusen, Bryan
; APPLICANT: Patunrajan, Meera
; APPLICANT: Shimkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangolli, Baha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Coleman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grose, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
```

```
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 258
; LENGTH: 662
; TYPE: PR
; ORGANISM: Homo sapiens
US-10-072-012-258
```

```
Query Match 45.7%; Score 304; DB 12; Length 662;
Best Local Similarity 100.0%; Pred. No. 1,1e-280;
Matches 304; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 362 SVPSVQPSLTPSPPIVQALSGHLSDADLEDSNKLKRFSLDIKVSYSASMAASLHGFS 421
|
|
|
Db 359 SVPSVQPSLTPSPPIVQALSGHLSDADLEDSNKLKRFSLDIKVSYSASMAASLHGFS 418
|
|
|
QY 422 SSEDALLEYKSTTLTGNTKLCQFSPVOELSEQTPEKSPKKEASIPKLTQTARPSDSQS 481
|
|
|
Db 419 SSEDALLEYKSTTLTGNTKLCQFSPVOELSEQTPEKSPKKEASIPKLTQTARPSDSQS 478
|
|
|
QY 482 KRLHSVRTSSSGTAQBSLSPLRSGSVEDNHTSFLFGLSTSQOHLTKSAGLKGWHS 541
|
|
|
Db 479 KRLHSVRTSSSGTAQBSLSPLRSGSVEDNHTSFLFGLSTSQOHLTKSAGLKGWHS 538
|
|
|
QY 542 DILAPQSTPSLTSSWYFATESHFPYSAAIYGSASYSASQSLPTCCDQVYSVRRQ 601
|
|
|
Db 539 DILAPQSTPSLTSSWYFATESHFPYSAAIYGSASYSASQSLPTCCDQVYSVRRQ 598
|
|
|
QY 602 KPSDRADSRSMHEESPFEKQFKRRSCOMEFGESIMSENRSREBELGKVGSSQSFSGSMET 661
|
|
|
Db 599 KPSDRADSRSMHEESPFEKQFKRRSCOMEFGESIMSENRSREBELGKVGSSQSFSGSMET 658
|
|
|
QY 662 IEVS 665
|
|
|
Db 659 IEVS 662
|
|
|
```

RESULT 18

```
US-10-296-115-1259
; Sequence 1259, Application US/10296115
; Publication No. US20040053248A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq Inc
; TITLE OF INVENTION: No. US20040053248A1 Nucleic Acids and Polypeptides
; FILE REFERENCE: 784PCT
; CURRENT APPLICATION NUMBER: US/10/296,115
; PRIOR FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: US09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US09/552,317
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 1478
; SEQ ID NO 1259
; LENGTH: 672
; TYPE: PR
; ORGANISM: Homo sapiens
; FEATRE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(672)
; OTHER INFORMATION: Xaa = any amino acid or other as shown in Table 3
US-10-296-115-1259
```

```
Query Match 38.8%; Score 258; DB 12; Length 672;
Best Local Similarity 100.0%; Pred. No. 7.8e-237;
Matches 258; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 194 TCPKPDFIPESHFLRVVUNVSFCCKILPWLDKSVDFLEKAKASGCVLHCLAGISSAT 253
|
|
|
Db 201 TCPKPDFIPESHFLRVVUNVSFCCKILPWLDKSVDFLEKAKASGCVLHCLAGISSAT 260
|
|
|
QY 254 IAIAYIMKRWDMSLDEAYRFEVKEKRPITSPNFPLGQLLDYERKKIKQTGASGPKSKLKL 313
|
|
|
```


/ LENGTH: 677
/ TYPE: PRT
/ ORGANISM: Mus musculus
US-10-072-012-683

Query Match 9.6%; Score 64; DB 12; Length 677;
Best Local Similarity 100.0%; Pred. No. 7.8e-52;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 114 SFSVHLGGFAPFCRCPGCGSTVPTCTISOPCLPVANIGRTILPNLYGCGORD 173
Db 114 SFSVHLGGFAPFCRCPGCGSTVPTCTISOPCLPVANIGRTILPNLYGCGORD 173

QY 174 VLNK 177
Db 174 VLNK 177

RESULT 21

US-10-072-012-702
/ Sequence 702, Application US/10072012
/ Publication No. US20040033493A1
/ GENERAL INFORMATION:

/ APPLICANT: Tchernev, Velizar
/ APPLICANT: Spytek, Kimberly
/ APPLICANT: Zernusen, Bryan
/ APPLICANT: Paturajan, Meera
/ APPLICANT: Shinkets, Richard
/ APPLICANT: Li, Li
/ APPLICANT: Gangolli, Esha
/ APPLICANT: Padigar, Muralidhara
/ APPLICANT: Anderson, David W.
/ APPLICANT: Rastelli, Luca
/ APPLICANT: Miller, Charles E.
/ APPLICANT: Gerlach, Valerie
/ APPLICANT: Taupier, Jr, Raymond J.
/ APPLICANT: Gusev, Vladimir Y.
/ APPLICANT: Coleman, Steven D.
/ APPLICANT: Wolenc, Adam R.
/ APPLICANT: Pena, Carol E. A
/ APPLICANT: Furtak, Katarzyna
/ APPLICANT: Grose, William M.
/ APPLICANT: Alsobrook II, John P.
/ APPLICANT: Lepley, Denise M.
/ APPLICANT: Rieger, Daniel K.
/ APPLICANT: Burgess, Catherine E.
/ TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
/ FILE REFERENCE: 21402-258
/ CURRENT APPLICATION NUMBER: US/10/072,012
/ CURRENT FILING DATE: 2002-01-31
/ PRIOR APPLICATION NUMBER: 60/265,102
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: 60/265,514
/ PRIOR FILING DATE: 2001-01-31
/ PRIOR APPLICATION NUMBER: 60/265,517
/ PRIOR FILING DATE: 2001-01-31
/ PRIOR APPLICATION NUMBER: 60/265,412
/ PRIOR FILING DATE: 2001-01-31
/ PRIOR APPLICATION NUMBER: 60/265,395
/ PRIOR FILING DATE: 2001-01-31
/ PRIOR APPLICATION NUMBER: 60/266,406
/ PRIOR FILING DATE: 2001-02-02
/ PRIOR APPLICATION NUMBER: 60/266,767
/ PRIOR FILING DATE: 2001-02-05
/ PRIOR APPLICATION NUMBER: 60/267,057
/ PRIOR FILING DATE: 2001-02-07
/ PRIOR APPLICATION NUMBER: 60/266,975
/ PRIOR FILING DATE: 2001-02-07
/ PRIOR APPLICATION NUMBER: 60/267,459
/ PRIOR FILING DATE: 2001-02-08
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 1391
/ SOFTWARE: PatentIn Ver. 2.1

/ SEQ ID NO 702
/ LENGTH: 501
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: VARIANT
/ LOCATION: (469)
/ OTHER INFORMATION: wherein Xaa is any amino acid.
US-10-072-012-702

Query Match 3.0%; Score 20; DB 12; Length 501;
Best Local Similarity 100.0%; Pred. No. 5.4e-10;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 242 VHCLAGISRSATIAIAYIMK 261
Db 98 VHCLAGISRSATIAIAYIMK 117

RESULT 22

US-10-072-012-699
/ Sequence 699, Application US/10072012
/ Publication No. US20040033493A1
/ GENERAL INFORMATION:

/ APPLICANT: Tchernev, Velizar
/ APPLICANT: Spytek, Kimberly
/ APPLICANT: Zernusen, Bryan
/ APPLICANT: Paturajan, Meera
/ APPLICANT: Shinkets, Richard
/ APPLICANT: Li, Li
/ APPLICANT: Gangolli, Esha
/ APPLICANT: Padigar, Muralidhara
/ APPLICANT: Anderson, David W.
/ APPLICANT: Rastelli, Luca
/ APPLICANT: Miller, Charles E.
/ APPLICANT: Gerlach, Valerie
/ APPLICANT: Taupier, Jr, Raymond J.
/ APPLICANT: Gusev, Vladimir Y.
/ APPLICANT: Coleman, Steven D.
/ APPLICANT: Wolenc, Adam R.
/ APPLICANT: Pena, Carol E. A
/ APPLICANT: Furtak, Katarzyna
/ APPLICANT: Grose, William M.
/ APPLICANT: Alsobrook II, John P.
/ APPLICANT: Lepley, Denise M.
/ APPLICANT: Rieger, Daniel K.
/ APPLICANT: Burgess, Catherine E.
/ TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
/ FILE REFERENCE: 21402-258
/ CURRENT APPLICATION NUMBER: US/10/072,012
/ CURRENT FILING DATE: 2002-01-31
/ PRIOR APPLICATION NUMBER: 60/265,102
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: 60/265,514
/ PRIOR FILING DATE: 2001-01-31
/ PRIOR APPLICATION NUMBER: 60/265,517
/ PRIOR FILING DATE: 2001-01-31
/ PRIOR APPLICATION NUMBER: 60/265,412
/ PRIOR FILING DATE: 2001-01-31
/ PRIOR APPLICATION NUMBER: 60/265,395
/ PRIOR FILING DATE: 2001-01-31
/ PRIOR APPLICATION NUMBER: 60/266,406
/ PRIOR FILING DATE: 2001-02-02
/ PRIOR APPLICATION NUMBER: 60/266,767
/ PRIOR FILING DATE: 2001-02-05
/ PRIOR APPLICATION NUMBER: 60/267,057
/ PRIOR FILING DATE: 2001-02-07
/ PRIOR APPLICATION NUMBER: 60/266,975
/ PRIOR FILING DATE: 2001-02-07
/ PRIOR APPLICATION NUMBER: 60/267,459
/ PRIOR FILING DATE: 2001-02-08
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 1391


```

; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 699
; LENGTH: 625
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-012-699

Query Match
Best Local Similarity 100.0%; Score 20; DB 12; Length 625;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 242 VHCLAGISRSATTAIAYIMK 261
DB 244 VHCLAGISRSATTAIAYIMK 263

RESULT 23
US-10-072-012-700
; Sequence 700, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:
; APPLICANT: Tchenev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zehrusen, Bryan
; APPLICANT: Patuturajan, Meera
; APPLICANT: Shinkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangoli, Esba
; APPLICANT: Padigaru, Muraidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie E.
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Colman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grosse, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 700
; LENGTH: 663
```

```

; TYPE: PRT
; ORGANISM: Mus musculus
US-10-072-012-700

Query Match
Best Local Similarity 100.0%; Score 20; DB 12; Length 663;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 242 VHCLAGISRSATTAIAYIMK 261
DB 244 VHCLAGISRSATTAIAYIMK 263

RESULT 24
US-10-072-012-266
; Sequence 266, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:
; APPLICANT: Tchenev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zehrusen, Bryan
; APPLICANT: Patuturajan, Meera
; APPLICANT: Shinkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangoli, Esba
; APPLICANT: Padigaru, Muraidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie E.
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Colman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grosse, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; PRIOR FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 266
; LENGTH: 616
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-072-012-266
```

Query Match 1.8%; Score 12; DB 12; Length 616;
Best Local Similarity 100.0%; Pred. No. 0.028;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 281 ISPNFNLGQL 292
|||||
DB 285 ISPNFNLGQL 296

RESULT 25
US-08-485-163-5
; Sequence 5, Application US/08485163
; Publication No. US2002098191A1
; GENERAL INFORMATION:
; APPLICANT: Beaudry, Gary A.
; TITLE OF INVENTION: CD4-GAMMA2 CD4-IGG2 CHIMERAS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,163
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 37690-II-1-PCT-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 530 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: homo sapien
; CELL TYPE: lymphocyte
US-08-485-163-5

Query Match 1.2%; Score 8; DB 8; Length 530;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 DRADSRRS 612
|||||
DB 78 DRADSRRS 85

RESULT 26
US-09-766-995-4
; Sequence 4, Application US/09766995
; Patent No. US20020052481A1
; GENERAL INFORMATION:
; APPLICANT: Graham P. Allaway et al.
; TITLE OF INVENTION: NON-PEPTIDYL MOIETY-CONJUGATED CD4-GAMMA2 AND CD4-IGG2 IMMUNOCONJUGATES
; FILE REFERENCE: 2048/41215-CB/JPM/SHS

; CURRENT APPLICATION NUMBER: US/09/766,995
; CURRENT FILING DATE: 2001-01-22
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: Patent in version 3.0
; SEQ ID NO: 4
; LENGTH: 530
; TYPE: PRT
; ORGANISM: homo sapiens

US-09-766-995-4

Query Match 1.2%; Score 8; DB 9; Length 530;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 DRADSRRS 612
|||||
DB 78 DRADSRRS 85

RESULT 27
US-09-939-537-6
; Sequence 6, Application US/09939537
; Publication No. US20030138410A1
; GENERAL INFORMATION:
; APPLICANT: Seed, Brian
; Banapour, Babak
; Romeo, Charles
; Kolanus, Waldemar
; TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED CELLS BY CHIMERIC CD4 RECEPTOR-BEARING CELLS
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Ebling LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/939,537
; FILING DATE: 24-Aug-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/284,391
; FILING DATE: 02-AUG-1994
; APPLICATION NUMBER: 08/195,395
; FILING DATE: 14-FEB-1994
; APPLICATION NUMBER: 07/847,566
; FILING DATE: 06-MAR-1992
; APPLICATION NUMBER: 07/665,961
; FILING DATE: 07-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Ebling, Karen L
; REGISTRATION NUMBER: 35,238
; REFERENCE/DOCKET NUMBER: 00786/247001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-428-0200
; TELEFAX: 617-428-7045
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 532 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-939-537-6

Query Match 1.2%: Score 8; DB 10; Length 532;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 DRADSRRS 612
|||||
Db 78 DRADSRRS 85

RESULT 28
US-09-243-008-6
; Sequence 6, Application US/09243008
; Publication No. US20040005334A1
; GENERAL INFORMATION:
; APPLICANT: Seed, Brian et al.
; TITLE OF INVENTION: Redirection of Cellular Immunity by
; Receptor Chimeras
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: Wordperfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/243,008
; FILING DATE: 02-Feb-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/394,176
; FILING DATE: SEPTEMBER 11, 1995
; APPLICATION NUMBER: 08/203,866
; FILING DATE: February 28, 1994
; APPLICATION NUMBER: 07/847,566
; FILING DATE: March 6, 1992
; APPLICATION NUMBER: 07/665,961
; FILING DATE: March 7, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Karen F. Lech, Ph.D.
; REGISTRATION NUMBER: 35,238
; REFERENCE/DOCKET NUMBER: 00786/270001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 532 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-243-008-6

Query Match 1.2%: Score 8; DB 11; Length 532;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 DRADSRRS 612
|||||
Db 78 DRADSRRS 85

RESULT 29
US-10-156-761-11387
; Sequence 11387, Application US/10156761
; Publication No. US20030119018A1

; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIGAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 11387
; LENGTH: 544
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-11387

Query Match 1.2%: Score 8; DB 14; Length 544;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 548 TSTPSLTS 555
|||||
Db 13 TSTPSLTS 20

RESULT 30
US-09-939-537-4
; Sequence 4, Application US/09939537
; Publication No. US20030138410A1
; GENERAL INFORMATION:
; APPLICANT: Seed, Brian
; Banapour, Babak
; Romeo, Charles
; Kolanus, Waldemar
; TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
; CELLS BY CHIMERIC CD4 RECEPTOR-BEARING CELLS
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Elbing LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/939,537
; FILING DATE: 24-Aug-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/284,391
; FILING DATE: 02-AUG-1994
; APPLICATION NUMBER: 08/195,395
; FILING DATE: 14-FEB-1994
; APPLICATION NUMBER: 07/847,566
; FILING DATE: 06-MAR-1992
; APPLICATION NUMBER: 07/665,961
; FILING DATE: 07-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Elbing, Karen L.
; REGISTRATION NUMBER: 35,238
; REFERENCE/DOCKET NUMBER: 00786/247001

TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617-428-0200
 TELEFAX: 617-428-7045
 TELEX: <unknown>
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 575 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 4:
 US-09-939-537-4

Query Match 1.2%; Score 8; DB 10; Length 575;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 605 DRADSRSS 612
 Db 78 DRADSRSS 85

RESULT 31
 US-09-243-008-4
 Sequence 4, Application US/09243008
 Publication No. US2004000534A1
 GENERAL INFORMATION:
 APPLICANT: Seed, Brian et al.
 TITLE OF INVENTION: Redirection of Cellular Immunity by Receptor Chimeras
 NUMBER OF SEQUENCES: 40
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fish & Richardson P.C.
 STREET: 225 Franklin Street
 CITY: Boston
 STATE: MA
 COUNTRY: USA
 ZIP: 02110-2804
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
 COMPUTER: IBM PS/2 Model 50Z or 55SX
 OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
 SOFTWARE: Wordperfect (Version 5.0)
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/243,008
 FILING DATE: 02-Feb-1999
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/08/394,176
 FILING DATE: SEPTEMBER 11, 1995
 APPLICATION NUMBER: 08/203,866
 FILING DATE: February 28, 1994
 APPLICATION NUMBER: 07/847,566
 FILING DATE: March 6, 1992
 APPLICATION NUMBER: 07/665,961
 FILING DATE: March 7, 1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Karen F. Lech, Ph.D.
 REGISTRATION NUMBER: 35,238
 REFERENCE/DOCKET NUMBER: 00786/270001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617) 542-5070
 TELEFAX: (617) 542-8906
 TELEX: 200154
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 575 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 4:
 US-09-243-008-4

Query Match 1.2%; Score 8; DB 11; Length 575;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 605 DRADSRSS 612
 Db 78 DRADSRSS 85

RESULT 32
 US-09-934-060A-13
 Sequence 13, Application US/09934060A
 Patent No. US20020155121A1
 GENERAL INFORMATION:
 APPLICANT: Devico, Anthony L.
 APPLICANT: Fouts, Timothy R.
 APPLICANT: Tuskan, Robert G.
 TITLE OF INVENTION: VIRUS COAT PROTEIN/RECEPTOR CHIMERAS AND METHODS OF USE
 FILE REFERENCE: 4115-144 CIP
 CURRENT APPLICATION NUMBER: US/09/934,060A
 CURRENT FILING DATE: 2001-08-21
 PRIOR APPLICATION NUMBER: US 09/684,026
 PRIOR FILING DATE: 2000-10-06
 PRIOR APPLICATION NUMBER: US 60/158,321
 PRIOR FILING DATE: 1999-10-08
 NUMBER OF SEQ ID NOS: 33
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 13
 LENGTH: 590
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Synthesized construct
 NAME/KEY: MISC FEATURE
 LOCATION: (586)..(586)
 OTHER INFORMATION: Xaa can be any amino acid
 NAME/KEY: MISC FEATURE
 LOCATION: (589)..(589)
 OTHER INFORMATION: Xaa can be any amino acid
 US-09-934-060A-13

Query Match 1.2%; Score 8; DB 9; Length 590;
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 605 DRADSRSS 612
 Db 450 DRADSRSS 457

RESULT 33
 US-09-934-060A-2
 Sequence 2, Application US/09934060A
 Patent No. US20020155121A1
 GENERAL INFORMATION:
 APPLICANT: Devico, Anthony L.
 APPLICANT: Fouts, Timothy R.
 APPLICANT: Tuskan, Robert G.
 TITLE OF INVENTION: VIRUS COAT PROTEIN/RECEPTOR CHIMERAS AND METHODS OF USE
 FILE REFERENCE: 4115-144 CIP
 CURRENT APPLICATION NUMBER: US/09/934,060A
 CURRENT FILING DATE: 2001-08-21
 PRIOR APPLICATION NUMBER: US 09/684,026
 PRIOR FILING DATE: 2000-10-06
 PRIOR APPLICATION NUMBER: US 60/158,321
 PRIOR FILING DATE: 1999-10-08
 NUMBER OF SEQ ID NOS: 33
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 2
 LENGTH: 720
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:

OTHER INFORMATION: Synthesized construct
NAME/KEY: MISC FEATURE
LOCATION: (716)..(716)
OTHER INFORMATION: Xaa can be any amino acid
NAME/KEY: MISC FEATURE
LOCATION: (719)..(719)
OTHER INFORMATION: Xaa can be any amino acid
US-09-934-060A-2

Query Match 1.2%; Score 8; DB 9; Length 720;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 DRADSRRS 612
DB 580 DRADSRRS 587

RESULT 34
US-09-934-060A-4
Sequence 4, Application US/09934060A
Patent No. US20020155121A1
GENERAL INFORMATION:
APPLICANT: DeVico, Anthony L.
APPLICANT: Fouts, Timothy R.
APPLICANT: Tusken, Robert G.
TITLE OF INVENTION: VIRUS COAT PROTEIN/RECEPTOR CHIMERAS AND METHODS OF USE
FILE REFERENCE: 4115-14 CIP
CURRENT APPLICATION NUMBER: US/09/934, 060A
PRIOR FILING DATE: 2001-08-21
PRIOR APPLICATION NUMBER: US 09/684, 026
PRIOR FILING DATE: 2000-10-06
PRIOR APPLICATION NUMBER: US 60/158, 321
PRIOR FILING DATE: 1999-10-08
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn version 3.1
SEQ ID NO 4
LENGTH: 720
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthesized construct
NAME/KEY: MISC FEATURE
LOCATION: (716)..(716)
OTHER INFORMATION: Xaa can be any amino acid
NAME/KEY: MISC FEATURE
LOCATION: (719)..(719)
OTHER INFORMATION: Xaa can be any amino acid
US-09-934-060A-4

Query Match 1.2%; Score 8; DB 9; Length 720;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 DRADSRRS 612
DB 580 DRADSRRS 587

RESULT 35
US-10-424-599-257796
Sequence 257796, Application US/10424599
Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: La Rosa Thomas J
APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
APPLICANT: Cao Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53223) B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28

NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 257796
LENGTH: 774
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(774)
OTHER INFORMATION: unsure at all Xaa locations
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_74813C.1.pep
US-10-424-599-257796

Query Match 1.2%; Score 8; DB 12; Length 774;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 647 GRVGSQSS 654
DB 736 GRVGSQSS 743

RESULT 36
US-10-073-118-26
Sequence 26, Application US/10073118
Publication No. US20030054554A1
GENERAL INFORMATION:
APPLICANT: BECQUART, JEROME
TITLE OF INVENTION: ALBUMIN DERIVATIVES WITH THERAPEUTIC FUNCTIONS
FILE REFERENCE: 06832.1429-03
CURRENT APPLICATION NUMBER: US/10/073,118
CURRENT FILING DATE: 2002-02-12
PRIOR APPLICATION NUMBER: 09/551,635
PRIOR FILING DATE: 2000-04-18
PRIOR APPLICATION NUMBER: 09/004,319
PRIOR FILING DATE: 1998-01-08
PRIOR APPLICATION NUMBER: 08/479,146
PRIOR FILING DATE: 1995-06-07
PRIOR APPLICATION NUMBER: 08/295,078
PRIOR FILING DATE: 1994-08-26
PRIOR APPLICATION NUMBER: 08/121,236
PRIOR FILING DATE: 1993-09-13
PRIOR APPLICATION NUMBER: 07/955,243
PRIOR FILING DATE: 1992-10-01
PRIOR APPLICATION NUMBER: 07/561,879
PRIOR FILING DATE: 1990-08-02
PRIOR APPLICATION NUMBER: FR 89 10480
PRIOR FILING DATE: 1989-08-03
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 26
LENGTH: 788
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Protein fusion
OTHER INFORMATION: prepro-HSA-VIIV2
US-10-073-118-26

Query Match 1.2%; Score 8; DB 14; Length 788;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 DRADSRRS 612
DB 662 DRADSRRS 669

RESULT 37
US-10-282-122A-47101
Sequence 47101, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:

```

; APPLICANT: Wang, Liangu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: EPIIIRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47101
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Borrelia burgdorferi
US-10-282-122A-47101

Query Match
Best Local Similarity 1.1%; Score 7; DB 12; Length 475;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 296 KXIKNOT 302
DB 469 KXIKNOT 475

RESULT 38
US-10-424-599-245600
; Sequence 245600, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 245600
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Glycine max
```

```

; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(478)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_63809C.1.pcp
US-10-424-599-245600

Query Match
Best Local Similarity 1.1%; Score 7; DB 12; Length 478;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 518 LFGLSTS 524
DB 125 LFGLSTS 131

RESULT 39
US-10-424-599-209874
; Sequence 209874, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 209874
; LENGTH: 479
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(479)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_31545C.1.pcp
US-10-424-599-209874

Query Match
Best Local Similarity 1.1%; Score 7; DB 12; Length 479;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 518 LFGLSTS 524
DB 126 LFGLSTS 132

RESULT 40
US-10-032-585-7589
; Sequence 7589, Application US/10032585
; Publication No. US20030180953A1
; GENERAL INFORMATION:
; APPLICANT: Terry, Roemer D.
; APPLICANT: Bo, Jiang
; APPLICANT: Charles, Boone
; APPLICANT: Howard, Bussey
; TITLE OF INVENTION: Gene Disruption Methodologies for Drug Target Discovery
; FILE REFERENCE: 10182-005-999
; CURRENT APPLICATION NUMBER: US/10/032,585
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 8000
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7589
; LENGTH: 498
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-032-585-7589
```

Query Match 1.1%; Score 7; DB 14; Length 498;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 489 TSSSGTA 495
|||||
DB 200 TSSSGTA 206

RESULT 41
US-10-425-114-62654
; Sequence 62654, Application US/10425114
; Publication No. US2004004888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaka, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; PRIOR FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 62654
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3066-034-Fl1.pep
US-10-425-114-62654

Query Match 1.1%; Score 7; DB 12; Length 503;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 229 FIREKAKA 235
|||||
DB 105 FIREKAKA 111

RESULT 42
US-10-355-238-13
; Sequence 13, Application US/10355238
; Publication No. US20040091987A1
; GENERAL INFORMATION:
; APPLICANT: SPELLIG, TILMAN
; TITLE OF INVENTION: PROCESS FOR THE OVEREXPRESSION OF DEHYDROGENASES
; FILE REFERENCE: SCH-1898
; CURRENT APPLICATION NUMBER: US/10/355,238
; PRIOR FILING DATE: 2003-01-31
; PRIOR APPLICATION NUMBER: DE 10204798.1
; PRIOR FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 60/382,569
; PRIOR FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 512
; TYPE: PRT
; ORGANISM: Brevibacterium maris
US-10-355-238-13

Query Match 1.1%; Score 7; DB 16; Length 512;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 424 EDALLEY 430
|||||
DB 72 EDALLEY 78

RESULT 43
US-10-355-238-23
; Sequence 23, Application US/10355238
; Publication No. US20040091987A1
; GENERAL INFORMATION:
; APPLICANT: SPELLIG, TILMAN
; TITLE OF INVENTION: PROCESS FOR THE OVEREXPRESSION OF DEHYDROGENASES
; FILE REFERENCE: SCH-1898
; CURRENT APPLICATION NUMBER: US/10/355,238
; PRIOR FILING DATE: 2003-01-31
; PRIOR APPLICATION NUMBER: DE 10204798.1
; PRIOR FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 60/382,569
; PRIOR FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 23
; LENGTH: 512
; TYPE: PRT
; ORGANISM: Brevibacterium maris
US-10-355-238-23

Query Match 1.1%; Score 7; DB 16; Length 512;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 424 EDALLEY 430
|||||
DB 72 EDALLEY 78

RESULT 44
US-10-355-238-14
; Sequence 14, Application US/10355238
; Publication No. US20040091987A1
; GENERAL INFORMATION:
; APPLICANT: SPELLIG, TILMAN
; TITLE OF INVENTION: PROCESS FOR THE OVEREXPRESSION OF DEHYDROGENASES
; FILE REFERENCE: SCH-1898
; CURRENT APPLICATION NUMBER: US/10/355,238
; PRIOR FILING DATE: 2003-01-31
; PRIOR APPLICATION NUMBER: DE 10204798.1
; PRIOR FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 60/382,569
; PRIOR FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Arthrobacter simplex
US-10-355-238-14

Query Match 1.1%; Score 7; DB 16; Length 515;
Best Local Similarity 100.0%; Pred. No. 1.4e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 424 EDALLEY 430
|||||
DB 71 EDALLEY 77

RESULT 45
US-10-355-238-25
; Sequence 25, Application US/10355238
; Publication No. US20040091987A1
; GENERAL INFORMATION:
; APPLICANT: SPELLIG, TILMAN
; TITLE OF INVENTION: PROCESS FOR THE OVEREXPRESSION OF DEHYDROGENASES
; FILE REFERENCE: SCH-1898
; CURRENT APPLICATION NUMBER: US/10/355,238


```

; CURRENT FILING DATE: 2003-01-31
; PRIOR APPLICATION NUMBER: DE 10204798.1
; PRIOR FILING DATE: 2002-02-01
; PRIOR APPLICATION NUMBER: 60/382,569
; PRIOR FILING DATE: 2002-05-24
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 25
; LENGTH: 515
; TYPE: PRT
; ORGANISM: Arthrobacter simplex
US-10-355-238-25

Query Match
Best Local Similarity 100.0%; Score 7; DB 16; Length 515;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 424 EDALLEY 430
DB 71 EDALLEY 77

RESULT 46
US-10-156-761-8798
; Sequence 8798, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8798
; LENGTH: 516
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-8798

Query Match
Best Local Similarity 100.0%; Score 7; DB 14; Length 516;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 225 KSVDFIE 231
DB 483 KSVDFIE 489

RESULT 47
US-10-072-012-732
; Sequence 732, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:
; APPLICANT: Tchernev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zernusen, Bryan
; APPLICANT: Patutajan, Meera
; APPLICANT: Shinkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangolli, Bsha
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
```

```

; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimir Y.
; APPLICANT: Coleman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grosse, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel K.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 732
; LENGTH: 523
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-072-012-732

Query Match
Best Local Similarity 100.0%; Score 7; DB 12; Length 523;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 488 RTSSSGT 494
DB 49 RTSSSGT 55

RESULT 48
US-10-282-122A-61404
; Sequence 61404, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zysek, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
```

```
FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61404
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Legionella pneumophila
US-10-282-122A-61404
```

```
Query Match      1.1%; Score 7; DB 12; Length 525;
Best Local Similarity 100.0%; Pred.No. 1.4e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      22 GTEKVL 28
        |||||
DB      225 GTEKVL 231
```

```
RESULT 49
US-10-361-460-73
; Sequence 73, Application US/10361460
; Publication NO. US20030163839A1
; GENERAL INFORMATION:
; APPLICANT: Helentjaris, Timothy G.
; APPLICANT: Bowen, Benjamin A.
; APPLICANT: Wang, Xun
; TITLE OF INVENTION: Genes Encoding Enzymes for Lignin
; TITLE OF INVENTION: Biosynthesis and Uses Thereof
; FILE REFERENCE: 0709
; CURRENT APPLICATION NUMBER: US/10/361,460
; CURRENT FILING DATE: 2003-02-10
; PRIOR APPLICATION NUMBER: 60/057,082
; PRIOR FILING DATE: 1997-08-27
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 73
; LENGTH: 526
; TYPE: PRT
; ORGANISM: Zea mays
US-10-361-460-73
```

```
Query Match      1.1%; Score 7; DB 14; Length 526;
Best Local Similarity 100.0%; Pred.No. 1.4e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      252 ATTAAIY 258
        |||||
DB      108 ATTAAIY 114
```

```
RESULT 50
US-10-289-757-106
; Sequence 106, Application US/10289757
; Publication NO. US20030180751A1
; GENERAL INFORMATION:
; APPLICANT: Demmer, Jeroen
; APPLICANT: Forester, Richard L
; APPLICANT: Gibson, John Bryan
; APPLICANT: Shenk, Michael Andrew
; APPLICANT: No. US20030180751A1r1s, Geoffrey
; APPLICANT: Glenn, Matthew
; APPLICANT: Hall, Claire
; APPLICANT: Saulsbury, Keith Martin
; TITLE OF INVENTION: Compositions isolated from forage
; TITLE OF INVENTION: grasses and methods for their use
; FILE REFERENCE: 11000.10610
; CURRENT APPLICATION NUMBER: US/10/289,757
; CURRENT FILING DATE: 2002-11-07
; PRIOR APPLICATION NUMBER: 60/337,703
; PRIOR FILING DATE: 2001-11-07
; NUMBER OF SEQ ID NOS: 218
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 106
; LENGTH: 532
; TYPE: PRT
; ORGANISM: Lolium perenne
US-10-289-757-106
```

```
Query Match      1.1%; Score 7; DB 14; Length 532;
Best Local Similarity 100.0%; Pred.No. 1.4e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      252 ATTAAIY 258
        |||||
DB      109 ATTAAIY 115
```

```
Search completed: June 21, 2004, 13:30:33
Job time : 50 secs
```

This Page Blank (uspto)

This Page Blank (uspto)

Wed Jun 23 06:36:37 2004

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: June 21, 2004, 13:25:49 ; Search time 22 Seconds
(without alignments)
1560.512 Million cell updates/sec

Title: US-10-029-345A-109

Perfect score: 665

Sequence: 1 MAHMEIGTIVETRLVALLE.....LKVSGSSFSGSMETIEVS 665

Scoring table: OLIGO
Gapex 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 28498

Minimum DB seq length: 473

Maximum DB seq length: 200000000

Post-processing: Listing first 75 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/prodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/prodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/2/iaa/backfillseq1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	472	71.0	665	4 US-09-816-494-2	Sequence 2, Appl1
2	8	1.2	530	3 US-08-477-460B-4	Sequence 4, Appl1
3	8	1.2	530	3 US-08-379-516-4	Sequence 4, Appl1
4	8	1.2	530	3 US-09-329-916-4	Sequence 4, Appl1
5	8	1.2	530	3 US-08-485-372A-4	Sequence 4, Appl1
6	8	1.2	530	4 US-09-409-006A-4	Sequence 4, Appl1
7	8	1.2	530	4 US-08-484-681-4	Sequence 4, Appl1
8	8	1.2	530	5 PCT-US93-07422-4	Sequence 4, Appl1
9	8	1.2	532	2 US-08-417-495-6	Sequence 6, Appl1
10	8	1.2	532	2 US-08-384-391B-6	Sequence 6, Appl1
11	8	1.2	532	3 US-09-218-950-6	Sequence 6, Appl1
12	8	1.2	532	5 PCT-US92-01785-6	Sequence 6, Appl1
13	8	1.2	532	5 PCT-US95-00454-6	Sequence 6, Appl1
14	8	1.2	575	2 US-08-417-495-4	Sequence 4, Appl1
15	8	1.2	575	2 US-08-384-391B-4	Sequence 4, Appl1
16	8	1.2	575	2 US-09-218-950-4	Sequence 4, Appl1
17	8	1.2	575	5 PCT-US92-01785-4	Sequence 4, Appl1
18	8	1.2	575	5 PCT-US95-00454-4	Sequence 4, Appl1
19	8	1.2	630	4 US-08-472-888A-6	Sequence 4, Appl1
20	8	1.2	903	1 US-08-021-601-12	Sequence 12, Appl1
21	8	1.2	903	1 US-08-082-849B-12	Sequence 12, Appl1
22	8	1.2	903	5 PCT-US94-01624-12	Sequence 12, Appl1
23	7	1.1	476	4 US-09-134-001C-4458	Sequence 4458, Ap
24	7	1.1	476	4 US-09-543-681A-7036	Sequence 7036, Ap
25	7	1.1	553	4 US-09-252-991A-32621	Sequence 32621, A
26	7	1.1	564	4 US-09-437-568A-2	Sequence 2, Appl1
27	7	1.1	569	1 US-08-306-231-3	Sequence 3, Appl1

ALIGNMENTS

RESULT 1
US-09-816-494-2
; Sequence 2, Application US/09816494
; Patent No. 6664089
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel A.
; TITLE OF INVENTION: 38692 AND 21117, NOVEL DUAL SPECIFICITY
; TITLE OF INVENTION: PHOSPHATASE MOLECULES AND USES THEREFOR
; FILE REFERENCE: 10448-030002
; CURRENT APPLICATION NUMBER: US/09/816,494
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 60/191,858
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 665
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-816-494-2

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/329,916
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/477,460
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: US 07/927,931
FILING DATE: 07-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 41215-A-PCT/JPW/AJM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 977-9809
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 530 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: homo sapien
CELL TYPE: lymphocyte
US-09-329-916-4

Query Match 1.2%; Score 8; DB 3; Length 530;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 605 DRADSRSS 612
Db 78 DRADSRSS 85

RESULT 5
US-08-485-372A-4
Sequence 4, Application US/08485372A
Patent No. 6187748
GENERAL INFORMATION:
APPLICANT: Beaudry, Gary A.
APPLICANT: Maddon, Paul J.
TITLE OF INVENTION: CD4-GAMMA2 CD4-IgG2 CHIMERAS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,372A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,227
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 37690-II-A

TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 530 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: homo sapien
CELL TYPE: lymphocyte
US-08-485-372A-4

Query Match 1.2%; Score 8; DB 3; Length 530;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 605 DRADSRSS 612
Db 78 DRADSRSS 85

RESULT 6
US-09-409-006A-4
Sequence 4, Application US/09409006A
Patent No. 6342586
GENERAL INFORMATION:
APPLICANT: Progenics Pharmaceuticals, Inc.
TITLE OF INVENTION: NON-PEPTIDYL MOIETY-CONJUGATED
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/409,006A
FILING DATE: 29-SEP-1999
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/927,931
FILING DATE: 07-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 41215-A-PCT/JPW/AJM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 977-9809
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 530 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: homo sapien
CELL TYPE: lymphocyte
US-09-409-006A-4

Query Match 1.2%; Score 8; DB 4; Length 530;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 DRADSRRS 612
|||||
DB 78 DRADSRRS 85

RESULT 7

US-08-484-681-4
Sequence 4, Application US/08484681
Patent No. 6451313
GENERAL INFORMATION:
APPLICANT: Beaudry, Gary A.
TITLE OF INVENTION: CD4-GAMMA2 CD4-IGG2 CHIMERAS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.24
CURRENT APPLICATION DATA:
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 37690-II-B
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 530 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: homo sapien
CELL TYPE: lymphocyte
US-08-484-681-4

Query Match 1.2%; Score 8; DB 4; Length 530;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 DRADSRRS 612
|||||
DB 78 DRADSRRS 85

RESULT 8
PCT-US93-07422-4
Sequence 4, Application PC/TUS9307422
GENERAL INFORMATION:
APPLICANT: Progenics Pharmaceuticals, Inc.
TITLE OF INVENTION: NON-PEPTIDYL MOIETY-CONJUGATED
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham

STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/07422
FILING DATE: 19930806
CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/927,931
FILING DATE: 07-AUG-1992
ATTORNEY/AGENT INFORMATION:

NAME: White, John P.
REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 41215-A-PCT/JPM/ALM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 977-9809

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:
LENGTH: 530 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:

ORGANISM: homo sapien
CELL TYPE: lymphocyte
PCT-US93-07422-4

Query Match 1.2%; Score 8; DB 5; Length 530;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 DRADSRRS 612
|||||
DB 78 DRADSRRS 85

RESULT 9

US-08-417-495-6
Sequence 6, Application US/08417495
Patent No. 5843728
GENERAL INFORMATION:
APPLICANT: Seed, Brian et al.
TITLE OF INVENTION: Redirection of Cellular Immunity by
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 50Z or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: Wordperfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/417,495
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/203,866

Chimeras

FILING DATE: US/07/847,566
APPLICATION NUMBER: 07/665,961
FILING DATE: March 7, 1991
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-417-495-6

Query Match 1.2%; Score 8; DB 2; Length 532;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 DRADSRSS 612
DB 78 DRADSRSS 85

RESULT 10

US-08-284-391B-6
Sequence 6, Application US/08284391B
Patent No. 5851828
GENERAL INFORMATION:
APPLICANT: Seed, Brian
APPLICANT: Banapour, Babak
APPLICANT: Romeo, Charles
APPLICANT: Kolanus, Waldemar
TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
CELLS BY CHIMERIC CD4 RECEPTOR-BEARING CELLS
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Clark & Elbing LLP
STREET: 176 Federal Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/284,391B
FILING DATE: 02-AUG-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/195,395
FILING DATE: 14-FEB-1994
APPLICATION NUMBER: 07/847,566
FILING DATE: 06-MAR-1992
APPLICATION NUMBER: 07/665,961
FILING DATE: 07-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Elbing, Karen L.
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/247001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045

TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-284-391B-6

Query Match 1.2%; Score 8; DB 2; Length 532;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 DRADSRSS 612
DB 78 DRADSRSS 85

RESULT 11

US-09-218-950-6
Sequence 6, Application US/09218950
Patent No. 6284240
GENERAL INFORMATION:
APPLICANT: Seed, Brian
APPLICANT: Banapour, Babak
APPLICANT: Romeo, Charles
APPLICANT: Kolanus, Waldemar
TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
CELLS BY CHIMERIC CD4 RECEPTOR-BEARING CELLS
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Clark & Elbing LLP
STREET: 176 Federal Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/218,950
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/284,391
FILING DATE: 02-AUG-1994
APPLICATION NUMBER: 08/195,395
FILING DATE: 14-FEB-1994
APPLICATION NUMBER: 07/847,566
FILING DATE: 06-MAR-1992
APPLICATION NUMBER: 07/665,961
FILING DATE: 07-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Elbing, Karen L.
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/247001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-218-950-6

Query Match 1.2%; Score 8; DB 3; Length 532;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 DRADSRSS 612
|||||||
Db 78 DRADSRSS 85

RESULT 12

PCT-US92-01785-6
; Sequence 6, Application PC/TUS9201785
; GENERAL INFORMATION:
; APPLICANT: The General Hospital Corporation
; TITLE OF INVENTION: Redirection of Cellular Immunity by Receptor
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: Wordperfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/01785
; FILING DATE: 19920306
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/665,961
; FILING DATE: March 7, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/119002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 532 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US92-01785-6

Query Match 1.2%; Score 8; DB 5; Length 532;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 DRADSRSS 612
|||||||
Db 78 DRADSRSS 85

RESULT 13

PCT-US95-00454-6
; Sequence 6, Application PC/TUS9500454
; GENERAL INFORMATION:
; APPLICANT: Seed, Brian et al.
; TITLE OF INVENTION: Targeted Cytolysis of HIV-Infected
; TITLE OF INVENTION: Cells by Chimeraic CD4 Receptor-
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street

CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: Wordperfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/00454
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/847,566
FILING DATE: March 6, 1992
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/665,961
FILING DATE: March 7, 1991
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/1247001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 532 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-00454-6

Query Match 1.2%; Score 8; DB 5; Length 532;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 DRADSRSS 612
|||||||
Db 78 DRADSRSS 85

RESULT 14
US-08-417-495-4
; Sequence 4, Application US/08417495
; Patent No. 5843728
; GENERAL INFORMATION:
; APPLICANT: Seed, Brian et al.
; TITLE OF INVENTION: Redirection of Cellular Immunity by
; TITLE OF INVENTION: Receptor
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: Wordperfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/417,495
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/203,866

FILING DATE:
APPLICATION NUMBER: US/07/847,566
FILING DATE:
APPLICATION NUMBER: 07/665,961
FILING DATE: March 7, 1991
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/119002
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 575 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-417-495-4

Query Match 1.2%; Score 8; DB 2; Length 575;
Best local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 DRADSRRS 612
DB 78 DRADSRRS 85

RESULT 15

US-08-284-391B-4
Sequence 4, Application US/08284391B
Patent No. 5851828
GENERAL INFORMATION:
APPLICANT: Seed, Brian
APPLICANT: Banapour, Babak
APPLICANT: Romeo, Charles
APPLICANT: Kolanus, Waldemar
TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
TITLE OF INVENTION: CELLS BY CHIMERIC CD4 RECEPTOR-BEARING CELLS
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Clark & Elding LLP
STREET: 176 Federal Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/284,391B
FILING DATE: 02-AUG-1994
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/195,395
FILING DATE: 14-FEB-1994
APPLICATION NUMBER: 07/847,566
FILING DATE: 06-MAR-1992
APPLICATION NUMBER: 07/665,961
FILING DATE: 07-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Elding, Karen L.
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/247001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045

TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 575 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-284-391B-4

Query Match 1.2%; Score 8; DB 2; Length 575;
Best local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 DRADSRRS 612
DB 78 DRADSRRS 85

RESULT 16

US-09-218-950-4
Sequence 4, Application US/09218950
Patent No. 6284240
GENERAL INFORMATION:
APPLICANT: Seed, Brian
APPLICANT: Banapour, Babak
APPLICANT: Romeo, Charles
APPLICANT: Kolanus, Waldemar
TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
TITLE OF INVENTION: CELLS BY CHIMERIC CD4 RECEPTOR-BEARING CELLS
NUMBER OF SEQUENCES: 53
CORRESPONDENCE ADDRESS:
ADDRESSEE: Clark & Elding LLP
STREET: 176 Federal Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/218,950
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/284,391
FILING DATE: 02-AUG-1994
APPLICATION NUMBER: 08/195,395
FILING DATE: 14-FEB-1994
APPLICATION NUMBER: 07/847,566
FILING DATE: 06-MAR-1992
APPLICATION NUMBER: 07/665,961
FILING DATE: 07-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Elding, Karen L.
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/247001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 575 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-218-950-4

Query Match 1.2%; Score 8; DB 3; Length 575;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 DRADSRSS 612
DB 78 DRADSRSS 85

RESULT 17

PCT-US92-01785-4
; Sequence 4, Application PC/TUS9201785
; GENERAL INFORMATION:
; APPLICANT: The General Hospital Corporation
; TITLE OF INVENTION: Redirection of Cellular Immunity by Receptor
; TITLE OF INVENTION: Chimeras
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 50Z or 55SX
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: Wordperfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/01785
; FILING DATE: 19920306
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/665,961
; FILING DATE: March 7, 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/119002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 575 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US92-01785-4

Query Match 1.2%; Score 8; DB 5; Length 575;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 DRADSRSS 612
DB 78 DRADSRSS 85

RESULT 18

PCT-US95-00454-4
; Sequence 4, Application PC/TUS9500454
; GENERAL INFORMATION:
; APPLICANT: Seed, Brian et al.
; TITLE OF INVENTION: Targeted Cytolysis of HIV-Infected
; TITLE OF INVENTION: Cells by Chimeric CD4 Receptor-
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street

CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 50Z or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: Wordperfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/00454
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/847,566
FILING DATE: March 6, 1992
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/665,961
FILING DATE: March 7, 1991
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/247001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 575 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-00454-4

Query Match 1.2%; Score 8; DB 5; Length 575;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 DRADSRSS 612
DB 78 DRADSRSS 85

RESULT 19

US-08-472-888A-6
; Sequence 6, Application US/08472888A
; Patent No. 6613746
; GENERAL INFORMATION:
; APPLICANT: Seed, Brian
; APPLICANT: Walz, Gerd
; TITLE OF INVENTION: AGP-ANTIBODY FUSION PROTEINS
; TITLE OF INVENTION: AND RELATED MOLECULES AND METHODS
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Erling LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,888A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/618,314
FILING DATE: 23-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Elbing, Karen L.
REGISTRATION NUMBER: 35,238
REFERENCE/DOCKET NUMBER: 00786/258001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-428-0200
TELEFAX: 617-428-7045
TELEX:
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 630 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-472-888A-6

Query Match 1.2%; Score 8; DB 4; Length 630;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 DRADSRSS 612
DB 78 DRADSRSS 85

RESULT 20
US-08-021-601-12
Sequence 12, Application US/08021601
Patent No. 5591631
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Nichols, Peter J.
APPLICANT: Aroora, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTIRAX TOXIN FUSION PROTEINS AND
TITLE OF INVENTION: RELATED METHODS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Needle & Rosenberg, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021,601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414.057
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 903 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-021-601-12

Query Match 1.2%; Score 8; DB 1; Length 903;

Best Local Similarity 100.0%; Pred. No. 44;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 DRADSRSS 612
DB 778 DRADSRSS 785

RESULT 21
US-08-082-849B-12
Sequence 12, Application US/08082849B
Patent No. 5677274
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Aroora, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
TITLE OF INVENTION: Related Methods
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 903 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-082-849B-12

Query Match 1.2%; Score 8; DB 1; Length 903;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 DRADSRSS 612
DB 778 DRADSRSS 785

RESULT 22
PCT-US94-01624-12
Sequence 12, Application PC/TUS9401624
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Aroora, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.

TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
TITLE OF INVENTION: RELATED METHODS
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESSES:
ADDRESSEE: TOWNSEND AND TOWNSEND KHOURIE AND CREW
STREET: Stewart Street Tower, 20th Floor, One Market
STREET: Plaza
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 903 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US94-01624-12

Query Match 1.2%; Score 8; DB 5; Length 903;
Best Local Similarity 100.0%; Pred. No. 44;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 605 DRADSRRS 612
DB 778 DRADSRRS 785

RESULT 23
US-09-134-001C-4458
Sequence 4458, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stramm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 4458
LENGTH: 476
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4458

Query Match 1.1%; Score 7; DB 4; Length 476;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 376 LVQALSG 382
DB 470 LVQALSG 476

RESULT 24
US-09-543-681A-7036
Sequence 7036, Application US/09543681A
Patent No. 6605709
GENERAL INFORMATION:
APPLICANT: GARY BRETON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543,681A
CURRENT FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US 60/128,706
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 7036
LENGTH: 486
TYPE: PRT
ORGANISM: Proteus mirabilis
US-09-543-681A-7036

Query Match 1.1%; Score 7; DB 4; Length 486;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 615 EESPFEX 621
DB 380 EESPFEX 386

RESULT 25
US-09-252-991A-32621
Sequence 32621, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 32621
LENGTH: 553
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32621

Query Match 1.1%; Score 7; DB 4; Length 553;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 349 AAGORPV 355
DB 152 AAGORPV 158

RESULT 26
US-09-437-568A-2
Sequence 2, Application US/09437568A
Patent No. 6620603
GENERAL INFORMATION:
APPLICANT: Lambeth, J. David
APPLICANT: Griendling, Kathy
APPLICANT: Lassegue, Bernard
APPLICANT: Arnold, Rebecca S.
APPLICANT: Cheng, Guangjie
TITLE OF INVENTION: NO. 6620603el Mitogenic Regulators

FILE REFERENCE: 05501-0103
CURRENT APPLICATION NUMBER: US/09/437,568A
CURRENT FILING DATE: 1999-11-10
NUMBER OF SEQ ID NOS: 61
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 2
LENGTH: 564
TYPE: PRT
ORGANISM: Homo sapiens
US-09-437-568A-2

Query Match 1.1%; Score 7; DB 4; Length 564;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 644 ELGKVG 650
DB 460 ELGKVG 466

RESULT 27

US-08-306-231-3
Sequence 3, Application US/08306231
Patent No. 5643748
GENERAL INFORMATION:
APPLICANT: Snodgrass, H. R.
APPLICANT: Ciofffi, Joseph
APPLICANT: Zupancic, Thomas J.
APPLICANT: Shafer, Alan W.
TITLE OF INVENTION: Hu-B1.219, A NOVEL HEMATOPOIETIN
TITLE OF INVENTION: RECEPTOR
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/306,231
FILING DATE: 14-SEP-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Polissant, Brian M.
REGISTRATION NUMBER: 28,462
REFERENCE/DOCKET NUMBER: 7225-076
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 569 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-306-231-3

Query Match 1.1%; Score 7; DB 1; Length 569;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 371 LEDSPV 377
DB 56 LEDSPV 62

RESULT 28

US-09-173-300-7
Sequence 7, Application US/09173300
Patent No. 6451581
GENERAL INFORMATION:
APPLICANT: Falco, Saverio Carl
APPLICANT: Hiltz, William D.
APPLICANT: Kinney, Anthony J.
APPLICANT: Cahoon, Rebecca E.
APPLICANT: Rafalski, J. Antoni
TITLE OF INVENTION: PLANT BRANCHED CHAIN AMINO ACID BIOSYNTHETIC ENZYMES
FILE REFERENCE: BB-1126
CURRENT APPLICATION NUMBER: US/09/173,300
CURRENT FILING DATE: 1998-10-15
EARLIER APPLICATION NUMBER: 60/063,423
EARLIER FILING DATE: 1997 October 28
NUMBER OF SEQ ID NOS: 54
SOFTWARE: Microsoft Word Version 7.0A
SEQ ID NO 7
LENGTH: 585
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-173-300-7

Query Match 1.1%; Score 7; DB 4; Length 585;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 235 ASNGCVL 241
DB 577 ASNGCVL 583

RESULT 29

US-09-252-991A-28887
Sequence 28887, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 28887
LENGTH: 679
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28887

Query Match 1.1%; Score 7; DB 4; Length 679;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 287 FLGQLLD 293
DB 397 FLGQLLD 403

RESULT 30

US-09-198-452A-848
Sequence 848, Application US/09198452A
Patent No. 6559294
GENERAL INFORMATION:
APPLICANT: Grifflais, R.
TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection

FILE REFERENCE: 9710-003-999
CURRENT APPLICATION NUMBER: US/09/198,452A
CURRENT FILING DATE: 1998-11-24
NUMBER OF SEQ ID NOS: 6849
SEQ ID NO 848
LENGTH: 687
TYPE: PRN
ORGANISM: Chlamydia pneumoniae
US-09-198-452A-848

Query Match 1.1%; Score 7; DB 4; Length 687;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 LIQHSK 74
DB 427 LIQHSK 433

RESULT 31
US-09-976-594-67
Sequence 67, Application US/09976594
Patent No. 6673549
GENERAL INFORMATION:
APPLICANT: Furness, Michael
APPLICANT: Buchbinder, Jenny
TITLE OF INVENTION: GENES EXPRESSED IN C3A LIVER CELL CULTURES TREATED WITH STEROIDS
FILE REFERENCE: PA-0041 US
CURRENT APPLICATION NUMBER: US/09/976,594
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: 60/240,409
PRIOR FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 1143
SOFTWARE: PERL Program
SEQ ID NO 67
LENGTH: 700
TYPE: PRN
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No. 6673549 3393396CD1
US-09-976-594-67

Query Match 1.1%; Score 7; DB 4; Length 700;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 CSQKVV 87
DB 479 CSQKVV 485

RESULT 32
US-08-533-306A-6
Sequence 6, Application US/0853306A
Patent No. 5837457
GENERAL INFORMATION:
APPLICANT: Liu, Pu
APPLICANT: Collins, Francis S.
APPLICANT: Siciliano, Michael J.
APPLICANT: Claxton, David
TITLE OF INVENTION: Markers for Detection of Chromosome 16
TITLE OF INVENTION: Rearrangements
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
STREET: P.O. Box 828
CITY: Bloomfield Hills
STATE: MI
COUNTRY: USA
ZIP: 48303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/533,306A
FILING DATE: September 25, 1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Smith, Deann F.
REGISTRATION NUMBER: 36683
REFERENCE/DOCKET NUMBER: 2115-00869COB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (810) 641-1600
TELEFAX: (810) 641-0270
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 816 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-533-306A-6

Query Match 1.1%; Score 7; DB 2; Length 816;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 DVASLS 100
DB 187 DVASLS 193

RESULT 33
US-08-742-923A-6
Sequence 6, Application US/08742923A
Patent No. 5869611
GENERAL INFORMATION:
APPLICANT: Liu, Pu
APPLICANT: Collins, Francis S.
APPLICANT: Siciliano, Michael J.
APPLICANT: Claxton, David
TITLE OF INVENTION: Markers for Detection of Chromosome 16
TITLE OF INVENTION: Rearrangements
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
STREET: P.O. Box 828
CITY: Bloomfield Hills
STATE: MI
COUNTRY: USA
ZIP: 48303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/742,923A
FILING DATE: No. 5869611ember 1, 1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Smith, Deann F.
REGISTRATION NUMBER: 36683
REFERENCE/DOCKET NUMBER: 2115-00869DVC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (810) 641-1600
TELEFAX: (810) 641-0270
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 816 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-742-923A-6

Query Match 1.1%; Score 7; DB 2; Length 816;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 DVASLSS 100
DB 187 DVASLSS 193

RESULT 34
US-09-134-000C-5599
Sequence 5599, Application US/09134000C
Patent No. 6617156
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stramm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
FILE REFERENCE: 032796-032
CURRENT APPLICATION NUMBER: US/09/134,000C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/055,778
PRIOR FILING DATE: 1997-08-15
NUMBER OF SEQ ID NOS: 6812
SOFTWARE: PatentIn version 3.1
SEQ ID NO 5599
LENGTH: 818
TYPE: PRT
ORGANISM: Enterococcus faecalis
US-09-134-000C-5599

Query Match 1.1%; Score 7; DB 4; Length 818;
Best Local Similarity 100.0%; Pred. No. 4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 510 EDNYHTS 516
DB 337 EDNYHTS 343

RESULT 35
US-08-533-306A-4
Sequence 4, Application US/08533306A
Patent No. 5837457
GENERAL INFORMATION:
APPLICANT: Liu, Pu
APPLICANT: Collins, Francis S.
APPLICANT: Siciliano, Michael J.
APPLICANT: Claxton, David
TITLE OF INVENTION: Markers for Detection of Chromosome 16
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSER: Harness, Dickey & Pierce, P.L.C.
STREET: P.O. Box 828
CITY: Bloomfield Hills
STATE: MI
COUNTRY: USA
ZIP: 48303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/533,306A
FILING DATE: September 25, 1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Smith, Deann F.
REGISTRATION NUMBER: 36683
REFERENCE/DOCKET NUMBER: 2115-00869COB
TELECOMMUNICATION INFORMATION:

TELEPHONE: (810) 641-1600
TELEFAX: (810) 641-0270
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 885 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-533-306A-4

Query Match 1.1%; Score 7; DB 2; Length 885;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 DVASLSS 100
DB 256 DVASLSS 262

RESULT 36
US-08-742-923A-4
Sequence 4, Application US/08742923A
Patent No. 5869611
GENERAL INFORMATION:
APPLICANT: Liu, Pu
APPLICANT: Collins, Francis S.
APPLICANT: Siciliano, Michael J.
APPLICANT: Claxton, David
TITLE OF INVENTION: Markers for Detection of Chromosome 16
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSER: Harness, Dickey & Pierce, P.L.C.
STREET: P.O. Box 828
CITY: Bloomfield Hills
STATE: MI
COUNTRY: USA
ZIP: 48303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/742,923A
FILING DATE: NO. 5869611ember 1, 1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Smith, Deann F.
REGISTRATION NUMBER: 36683
REFERENCE/DOCKET NUMBER: 2115-00869DVC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (810) 641-1600
TELEFAX: (810) 641-0270
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 885 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-742-923A-4

Query Match 1.1%; Score 7; DB 2; Length 885;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 DVASLSS 100
DB 256 DVASLSS 262

RESULT 37
US-09-540-236-2916

; Sequence 2916, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540.236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 2916
; LENGTH: 888
; TYPE: PRT
; ORGANISM: M. catarrhalis
US-09-540-236-2916

Query Match 1.1%; Score 7; DB 4; Length 888;
Best Local Similarity 100.0%; Pred. No. 4.3e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 10 IYTERLV 16
Db 552 IYTERLV 558

RESULT 38
US-08-640-389A-10
; Sequence 10, Application US/08640389A
; Patent No. 5912123
; GENERAL INFORMATION:
; APPLICANT: Snodgrass, H. R.
; APPLICANT: Cioffi, Joseph
; APPLICANT: Zupancic, Thomas J.
; APPLICANT: Shafer, Alan W.
; TITLE OF INVENTION: DETECTION OF THE LEPTIN
; TITLE OF INVENTION: RECEPTOR IN REPRODUCTIVE ORGANS AND METHODS FOR
; TITLE OF INVENTION: REGULATING REPRODUCTIVE BIOLOGY
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/640,389A
; FILING DATE: 29-APR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 8907-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 896 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-640-389A-10

Query Match 1.1%; Score 7; DB 2; Length 896;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 371 LEDSPV 377
Db 169 LEDSPV 175

RESULT 39
US-08-618-957A-10
; Sequence 10, Application US/08618957A
; Patent No. 6355237
; GENERAL INFORMATION:
; APPLICANT: Snodgrass, H. Ralph
; APPLICANT: Cioffi, Joseph
; APPLICANT: Zupancic, Thomas Joel
; APPLICANT: Shafer, Alan Wayne
; TITLE OF INVENTION: METHODS FOR USING THE OBES
; TITLE OF INVENTION: GENE AND ITS GENE PRODUCT TO STIMULATE HEMATOPOIETIC
; TITLE OF INVENTION: DEVELOPMENT
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Pennie & Edmonds LLP
; STREET: 1155 Avenue of The Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/618,957A
; FILING DATE: 20-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 008907-0033-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 896 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-618-957A-10

Query Match 1.1%; Score 7; DB 4; Length 896;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 371 LEDSPV 377
Db 169 LEDSPV 175

RESULT 40
US-09-043-816E-13
; Sequence 13, Application US/09043816E
; Patent No. 6414128
; GENERAL INFORMATION:
; APPLICANT: Hilton, Douglas J.
; APPLICANT: Willson, Tracy
; APPLICANT: Nicola, Nicos A.

```

; APPLICANT: Gainsford, Timothy
; APPLICANT: Alexander, Warren S.
; APPLICANT: Metcalf, Donald
; APPLICANT: Ng, Ashley
; TITLE OF INVENTION: A NOVEL HAEMOPOIETIN RECEPTOR AND GENETIC SEQUENCES
; FILE REFERENCE: 11268
; CURRENT APPLICATION NUMBER: US/09/043,816E
; CURRENT FILING DATE: 1998-09-17
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 13
; LENGTH: 896
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (223)
; OTHER INFORMATION: Xaa is unknown or other.
; NAME/KEY: UNSURE
; LOCATION: (687)
; OTHER INFORMATION: Xaa is unknown or other.
; US-09-043-816E-13
```

```

Query Match      1.1%; Score 7; DB 4; Length 896;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      371 LEDSPV 377
        |||||
Db      169 LEDSPV 175
```

```

RESULT 41
; US-09-357-914-33
; Sequence 33, Application US/09357914
; Patent No. 6524806
; GENERAL INFORMATION:
; APPLICANT: Snodgrass, H. Ralph
; APPLICANT: Cioffi, Joseph
; APPLICANT: Zupancic, Thomas J.
; APPLICANT: Shafer, Alan Wayne
; TITLE OF INVENTION: ANTIBODIES SPECIFIC FOR Hu-B1.219, A
; FILE REFERENCE: 8907-0083-999
; CURRENT APPLICATION NUMBER: US/09/357,914
; CURRENT FILING DATE: 1999-07-19
; PRIOR APPLICATION NUMBER: US 08/693,696
; PRIOR FILING DATE: 1996-08-05
; PRIOR APPLICATION NUMBER: US 08/355,888
; PRIOR FILING DATE: 1994-12-14
; PRIOR APPLICATION NUMBER: US 08/306,231
; PRIOR FILING DATE: 1994-09-14
; NUMBER OF SEQ. ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33
; LENGTH: 896
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-357-914-33
```

```

Query Match      1.1%; Score 7; DB 4; Length 896;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      371 LEDSPV 377
        |||||
Db      169 LEDSPV 175
```

```

RESULT 42
; US-08-780-562-3
; Sequence 3, Application US/08780562
```

```

; Patent No. 6541604
; GENERAL INFORMATION:
; APPLICANT: Matthews, William
; APPLICANT: Bennett, Brian
; TITLE OF INVENTION: MSX RECEPTOR
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/780,562
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/585005
; FILING DATE: 01/08/97
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/
; FILING DATE: 01/08/97
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee, Wendy M.
; REGISTRATION NUMBER: 40,378
; REFERENCE/DOCKET NUMBER: P0986R1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1994
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 896 amino acids
; TYPE: Amino Acid
; TOPOLOGY: Linear
; US-08-780-562-3
```

```

Query Match      1.1%; Score 7; DB 4; Length 896;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      371 LEDSPV 377
        |||||
Db      169 LEDSPV 175
```

```

RESULT 43
; US-08-693-697-36
; Sequence 36, Application US/08693697
; Patent No. 5869610
; GENERAL INFORMATION:
; APPLICANT: Snodgrass, H. R.
; APPLICANT: Cioffi, Joseph
; APPLICANT: Zupancic, Thomas J.
; APPLICANT: Shafer, Alan W.
; TITLE OF INVENTION: Hu-B1.219, A NOVEL HUMAN HAEMOPOIETIN
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Penite & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/693,697
FILING DATE: 05-AUG-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Poissant, Brian M.
REGISTRATION NUMBER: 28,462
REFERENCE/DOCKET NUMBER: 8907-0037-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-493-4935
TELEFAX: 650-493-5556
TELEX: 66141 PENNITE
INFORMATION FOR SEQ ID NO: 36:
SEQUENCE CHARACTERISTICS:
LENGTH: 898 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-693-697-36

Query Match
Best Local Similarity 1.1%; Score 7; DB 2; Length 898;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 371 LEDSPV 377
DB 171 LEDSPV 177

RESULT 44
US-08-588-189-3
Sequence 3, Application US/08588189
Patent No. 6451523
GENERAL INFORMATION:
APPLICANT: Snodgrass, H.
APPLICANT: Cioffi, Joseph
APPLICANT: Zupancic, Thomas
APPLICANT: Shafer, Alan
TITLE OF INVENTION: DETECTION OF A LEPTIN RECEPTOR VARIANT
TITLE OF INVENTION: AND METHODS FOR REGULATING OBESITY
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: US
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/588,189
FILING DATE: 18-JAN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Poissant, Brian M.
REGISTRATION NUMBER: 28,462
REFERENCE/DOCKET NUMBER: 8907-031
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741
TELEX: 66141 PENNITE
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 898 amino acids

TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-588-189-3

Query Match
Best Local Similarity 1.1%; Score 7; DB 4; Length 898;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 371 LEDSPV 377
DB 171 LEDSPV 177

RESULT 45
US-08-640-389A-9
Sequence 9, Application US/08640389A
Patent No. 5912123
GENERAL INFORMATION:
APPLICANT: Snodgrass, H. R.
APPLICANT: Cioffi, Joseph
APPLICANT: Zupancic, Thomas J.
APPLICANT: Shafer, Alan W.
TITLE OF INVENTION: DETECTION OF THE LEPTIN
TITLE OF INVENTION: RECEPTOR IN REPRODUCTIVE ORGANS AND METHODS FOR
TITLE OF INVENTION: REGULATING REPRODUCTIVE BIOLOGY
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/640,389A
FILING DATE: 29-APR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Poissant, Brian M.
REGISTRATION NUMBER: 28,462
REFERENCE/DOCKET NUMBER: 8907-032
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNITE
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 906 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-640-389A-9

Query Match
Best Local Similarity 1.1%; Score 7; DB 2; Length 906;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 371 LEDSPV 377
DB 169 LEDSPV 175

RESULT 46
US-08-618-957A-9
Sequence 9, Application US/08618957A
Patent No. 6355237

```

; GENERAL INFORMATION:
; APPLICANT: Snodgrass, H. Ralph
; APPLICANT: Cioffi, Joseph
; APPLICANT: Zupancic, Thomas Joel
; APPLICANT: Shafer, Alan Wayne
; TITLE OF INVENTION: METHODS FOR USING THE OBSE
; TITLE OF INVENTION: GENE AND ITS GENE PRODUCT TO STIMULATE HEMATOPOIETIC
; TITLE OF INVENTION: DEVELOPMENT
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of The Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/618,957A
; FILING DATE: 20-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 008907-0033-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 906 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-618-957A-9

Query Match 1.1%; Score 7; DB 4; Length 906;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 371 LEDSPV 377
DB 169 LEDSPV 175

RESULT 47
US-09-357-914-32
; Sequence 32, Application US/09357914
; Patent No. 6524806
; GENERAL INFORMATION:
; APPLICANT: Snodgrass, H. Ralph
; APPLICANT: Cioffi, Joseph
; APPLICANT: Zupancic, Thomas J.
; APPLICANT: Shafer, Alan Wayne
; TITLE OF INVENTION: ANTIBODIES SPECIFIC FOR Hu-B1.219, A
; TITLE OF INVENTION: NOVEL HUMAN HEMATOPOIETIN RECEPTOR
; FILE REFERENCE: 8907-0083-999
; CURRENT APPLICATION NUMBER: US/09/357,914
; CURRENT FILING DATE: 1999-07-19
; PRIOR APPLICATION NUMBER: US 08/693,696
; PRIOR FILING DATE: 1996-08-05
; PRIOR APPLICATION NUMBER: US 08/355,888
; PRIOR FILING DATE: 1994-12-14
; PRIOR APPLICATION NUMBER: US 08/306,231
```

```

; PRIOR FILING DATE: 1994-09-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 32
; LENGTH: 906
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-357-914-32

Query Match 1.1%; Score 7; DB 4; Length 906;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 371 LEDSPV 377
DB 169 LEDSPV 175

RESULT 48
US-08-693-697-33
; Sequence 33, Application US/08693697
; Patent No. 5869610
; GENERAL INFORMATION:
; APPLICANT: Snodgrass, H. R.
; APPLICANT: Cioffi, Joseph
; APPLICANT: Zupancic, Thomas J.
; APPLICANT: Shafer, Alan W.
; TITLE OF INVENTION: Hu-B1.219, A NOVEL HUMAN HEMATOPOIETIN
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of The Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/693,697
; FILING DATE: 03-AUG-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 8907-0037-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 908 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-08-693-697-33

Query Match 1.1%; Score 7; DB 2; Length 908;
Best Local Similarity 100.0%; Pred. No. 4.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 371 LEDSPV 377
DB 171 LEDSPV 177
```

RESULT 49

US-08-588-526-3

; Sequence 3, Application US/08588526
; Patent No. 5882860

; GENERAL INFORMATION:

; APPLICANT: Snodgrass, H.
; APPLICANT: Cioffi, Joseph
; APPLICANT: Zupancic, Thomas
; APPLICANT: Shafer, Alan
; TITLE OF INVENTION: DETECTION OF A LEPTIN RECEPTOR
; TITLE OF INVENTION: VARIANT
; TITLE OF INVENTION: AND METHODS FOR REGULATING OBESITY
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: US
; ZIP: 10036-2711

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/588,526
; FILING DATE: 18-JAN-1996
; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Poissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 8907-030
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741

; TELEX: 66141 PENNIE

; INFORMATION FOR SEQ. ID NO: 3:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 908 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-588-526-3

Query Match 1.1%; Score 7; DB 2; Length 908;

Best Local Similarity 100.0%; Pred. No. 4.4e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 371 LEDSPV 377
Db 171 LEDSPV 177

RESULT 50

US-08-780-562-4

; Sequence 4, Application US/08780562
; Patent No. 6541604

; GENERAL INFORMATION:

; APPLICANT: Matthews, William
; APPLICANT: Bennett, Brian
; TITLE OF INVENTION: WSX RECEPTOR
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
; COMPUTER: IBM PC compatible; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WinPatIn (Genentech); CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/780,562

; FILING DATE:

; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/585005
; FILING DATE: 01/08/97
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/
; FILING DATE: 01/08/97
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee, Wendy M.
; REGISTRATION NUMBER: 40,378
; REFERENCE/DOCKET NUMBER: P0986R1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415/225-1994
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ. ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 923 amino acids
; TYPE: Amino Acid
; TOPOLOGY: linear
; US-08-780-562-4

Query Match

1.1%; Score 7; DB 4; Length 923;

Best Local Similarity 100.0%; Pred. No. 4.5e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 371 LEDSPV 377
Db 169 LEDSPV 175Search completed: June 21, 2004, 13:29:32
Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 22, 2004, 05:14:42 ; Search time 6973 Seconds

(without alignments)
4133.532 Million cell updates/sec

Title: US-10-029-345a-109

Perfect score: 665
Sequence: 1 MAHEMIGTIVTERLVALL.....LGVSGSSFSGSMELIEVS 665

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 2167151695 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1263448

Minimum DB seq length: 1419

Maximum DB seq length: 200000000

Post-processing: Listing first 75 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlh
-C/cgn2.1/USPTO.spool/US1002345/runat.21062004.164401.29751/bpp.query.fasta.1.839
-DB=GenEmbl -QFMT=fastap -SUFFIX=oligo.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=75
-DOCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=50 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZE=500 -MINLEN=1419 -MAXLEN=200000000
-USER=US1002345_@CGN_1_1_4290_@runat.21062004.164401.29751 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl:.*
1: gb_ba:.*
2: gb_hcg:.*
3: gb_in:.*
4: gb_om:.*
5: gb_ov:.*
6: gb_pat:.*
7: gb_ph:.*
8: gb_pl:.*
9: gb_pr:.*
10: gb_ro:.*
11: gb_sts:.*
12: gb_sy:.*
13: gb_un:.*
14: gb_vi:.*
15: em_ba:.*
16: em_fun:.*
17: em_hum:.*
18: em_in:.*
19: em_mu:.*
20: em_om:.*
21: em_or:.*
22: em_ov:.*
23: em_pat:.*
24: em_ph:.*
25: em_pl:.*
26: em_ro:.*
27: em_sts:.*
28: em_un:.*

29: em_vi:.*
30: em_hcg_hum:.*
31: em_hcg_inv:.*
32: em_hcg_other:.*
33: em_hcg_mus:.*
34: em_hcg_pln:.*
35: em_hcg_rod:.*
36: em_hcg_mam:.*
37: em_hcg_vrt:.*
38: em_sy:.*
39: em_hcgo_hum:.*
40: em_hcgo_mus:.*
41: em_hcgo_other:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	665	100.0	5450	6	AX482439 Sequence
2	665	100.0	5450	6	AX482478 Sequence
3	472	71.0	1935	9	AY038927 Homo sapi
4	472	71.0	1998	6	AX260342 Sequence
5	472	71.0	2102	6	AX713989 Sequence
6	472	71.0	2102	9	AK055973 Homo sapi
7	472	71.0	2732	6	AX180875 Sequence
8	472	71.0	3059	6	AX278461 Sequence
9	472	71.0	3104	6	AX405700 Sequence
10	472	71.0	3332	6	AX441229 Sequence
11	472	71.0	3496	6	AX441210 Sequence
12	472	71.0	3521	9	AB052156 Homo sapi
13	472	71.0	3544	6	AX260340 Sequence
14	472	71.0	3566	9	AF506796 Homo sapi
15	472	71.0	3766	6	AX374994 Sequence
16	472	71.0	4790	6	BD171157 Novel gen
17	472	71.0	4790	6	BD183422 Novel gen
18	472	71.0	4790	9	AB051487 Homo sapi
19	472	71.0	4790	6	AB051487 Homo sapi
20	472	71.0	4790	6	AB051487 Homo sapi
21	472	71.0	4790	6	AB051487 Homo sapi
22	472	71.0	4790	6	AB051487 Homo sapi
23	472	71.0	4790	6	AB051487 Homo sapi
24	472	71.0	4790	6	AB051487 Homo sapi
25	472	71.0	4790	6	AB051487 Homo sapi
26	472	71.0	4790	6	AB051487 Homo sapi
27	472	71.0	4790	6	AB051487 Homo sapi
28	472	71.0	4790	6	AB051487 Homo sapi
29	472	71.0	4790	6	AB051487 Homo sapi
30	472	71.0	4790	6	AB051487 Homo sapi
31	472	71.0	4790	6	AB051487 Homo sapi
32	472	71.0	4790	6	AB051487 Homo sapi
33	472	71.0	4790	6	AB051487 Homo sapi
34	472	71.0	4790	6	AB051487 Homo sapi
35	472	71.0	4790	6	AB051487 Homo sapi
36	472	71.0	4790	6	AB051487 Homo sapi
37	472	71.0	4790	6	AB051487 Homo sapi
38	472	71.0	4790	6	AB051487 Homo sapi
39	472	71.0	4790	6	AB051487 Homo sapi
40	472	71.0	4790	6	AB051487 Homo sapi
41	472	71.0	4790	6	AB051487 Homo sapi

OY	321	GIUPROVALP	PROVALA	VALSERGLUGLYGLYGLINLYS	SERGLUTHP	PROLEUSER	PRO	340
Db	1498	GAACCTGTC	CCCTGCTCTCAGAGGGTGGAGAAA	AAACGAGAGCCCTTCA	GTCCACCC			
OY	341	CYAA	AAAPSERALAT	ThrSerGluAla	AlaGlyGln	ArgProVal	His	360
Db	1558	TGTGCCGACT	CTTGTA	CTCTCAGAGGCACAGACAA	AAAGCCCGTG	CACTCCGCA	GCGTG	1617
OY	361	PROSERVAL	PROSERVAL	GlnProSerLeu	GluAspSer	ProLeuVal	GlnAla	380
Db	1618	CCGACGCGGCC	AGCGTGCCAGCCGCTGTT	AGAGGACACACCCGCTG	TGTAACAGCGCTC			1677
OY	381	SERGLYLEU	His	LEUSERAL	AspArgLeu	GluAspSer	Anly	400
Db	1678	AGTGGCGTCA	CCCTGTCCCGACAGACAGCGTGA	AGACAAATTAAGCTCA	AAACGTTCCCTTC			1737
OY	401	SERLEUAP	PII	ELYSERVAL	SerTyrSer	AlaSerMet	Ala	420
Db	1738	TCTCGGATATCA	AAATCA	ATTTCATATTCACCCAGCAT	GGCAGCATCTTACATG	GGCTTC		1797
OY	421	SER	SER	SERGLI	AspAla	LeuGlyI	uTyrTyr	440
Db	1798	TCTCATGAGA	NAAGNGCTTGG	AAATCAACAACCTTCC	ATCTCTGATGGG	ACCAAC		1857
OY	441	LYS	LEU	CYSGln	His	SER	ProVal	460
Db	1858	AAAGCTATCC	AGATTCTCCCTGTT	CAGAACTATG	CGAGACACTCCG	AAACAGTCTT		1917
OY	461	AS	PLY	SGI	UGI	UNAL	SER	480
Db	1918	GATA	AGGAGGAAGCCAG	CATCCCAAGAGCTCCAG	ACCGGACAGCTTCA	GACAGCCAG		1977
OY	481	SER	LY	SA	AGLY	LEU	His	500
Db	1978	AGCA	AGC	ATTGCA	CTTCGCTCAG	AAACCAAGCAG	CAGTGGCACCGCCAG	2037
OY	501	SER	PRO	LEU	His	ARG	SER	520
Db	2038	TCTCCACT	GCATCGAA	GTGGAGCGTGGAGGA	CAATTA	CCACA	CCAGCTTCTTTT	2097
OY	521	LEU	SER	THR	SER	GlnGln	His	540
Db	2098	CTTTCAC	CACGACGAC	CACACCTCA	CGAAGTCTGTG	CCCTTA	AAAGGCTGG	2157
OY	541	SER	AS	PII	LEU	Ala	Pro	560
Db	2158	TGCA	TATCTTG	GGCCCCCAG	ACCTCTAC	CCCTTCC	TGACACGAC	2217
OY	561	THR	Gln	SER	SER	His	Ala	580
Db	2218	ACAG	AGTCTCTCA	CTTCTACT	CTGCTCA	GCACATCTA	CGAGGAGTCA	2277
OY	581	ALA	Tyr	SER	CY	SER	Gln	600
Db	2278	GCCT	TCACACTG	AGCCAGCAGCTG	CCCACTTGC	CGAGACCA	AAGTCTATTCTG	2337
OY	601	Gln	LY	SER	PRO	SER	Asp	620
Db	2338	CAGA	AGCCAA	GTGAAGAGCTG	ACTCGCGGCG	AGCTGCA	CTGAAGAAAGCCCTT	2397
OY	621	LYS	Gln	PII	Leu	Arg	SER	640
Db	2398	AA	GCAGTTTAA	ACGCAAGAC	GTGCCAAAT	TGGAAATTG	GAGAGCATCATC	2457
OY	641	ARG	SER	ARG	Gln	Leu	Gly	660
Db	2458	AGGT	ACACGG	AGAGACTG	CGGGAAGATGG	GCAGTCA	GTCTTTTCGG	2517
OY	661	ILE	ILE	Gln	Val	SER	665	
Db	2518	ATCAT	TGAGGTCTCC	2532				

AX482478 AX482478 5450 bp DNA linear PAT 16-AUG-2002
LOCUS Sequence 147 from Patent WO02057460.
DEFINITION AX482478
ACCESSION AX482478
VERSION AX482478.1 GI:22316999
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS 1
TODDERUD,C.G., BOL,D., FINGER,J., LEE,L., NELSON,T., SCHLIVEN,G.,
SUCHARD,S., BANS,D., BAASSOLINO,D., PEDER,J., KRYSTEK,S.,
MCATEE,P., MINTLER,G., SIEMERS,N., JACKSON,D.G. and RAMANATHAN,C.
Polynucleotides encoding human phosphatases
Patent: WO 02057460-A 147 25 -JUL-2002;
JOURNAL SQUIBB BRISTOL MYERS CO (US)
FEATURES
Source Location/Qualifiers
1..5450
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
538..2535
/notes="unamed protein product"
/codon_start=1
/protein_id="CAD44453.1"
/db_xref="GI:22317000"
/db_xref="REMBL:CAD44453"
/translation="MAHEMIGTQIVTERIALVLESGTEKYLIDSRPFVEYNTSHILE
AININCSKLMKRLQDQKVLTELILQHSARKKVIDDSQKRVYVDSSQVYASSSDPC
FLTVLGLGKLEKSFNSVHLAAGFAEFRCPCFGEKSTLPTISQCLPVAINGP
RLTPLYLQCGORDYANKELIQONGIGVLANSYCPKPDFIPESHFLRPVNDPCEKE
ILPMLDKSVDFIEKAKASNGCVLNVHCLAGLSRSATIIAYIMKMDMSLDGAYFVKE
KRPITSPENFNGQLDLYEKKIKNOTGSGSKTLKLIHEKPEYPAVSEGGOKS
TLPSPCADSATSEAAAGORPHPASVSVPVQSLLEDSPLNALGILNLSARLEKE
SNKRSRSDIDKSVYSASMAASLHGFSESDLEBYKEPTTIDGTRNLQCSFVQV
LSEDTPEPSPDEKASIPKQLQTPRPDSQSRHVSRTSSGTAQNSLSPHRSGS
VEDNYHSLFGLSTSOHLTKSAGLKGWHSOILAPQTPSLTSWYPATSSHQP
YSAAIYGSGASVAYSCSLPTCGDPDQVSVRRQKPSDRADSRSSWHEPFEKQ
RRSQMERGESIMSENRSELGKVGSGSSPFGSMELIEVS"

ORIGIN

Alignment Scores:

Pred. No.:	0	Length:	5450
Score:	665.00	Matches:	665
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-10-029-345A-109 (1-665) x AX482478 (1-5450)

QY 1 MetAlaHlsglUmEtllEgLYThrglnlEvAlThrgluAglEuvAlAlAlEulEuglu 20

Db 538 ATGGCCCATGAGATGATGGAACCAATTGTTACTGAGAGTTGGCGCTCTCTGGAA 597

QY 21 SerGIYThrgluYvAlEulEulEulEkePseArPrPopeVAlAGluTYAsnThsSer 40

Db 538 AGTGAACGGAAGAAAGTCTCTAATGATATGATGCGGCGCATTTGGTGAGTACATATCATCC 657

QY 41 HislELeuglAlAlAlEAsnllEAsnCYseSerlySleuMetlyBArGArgLeuglnG 60

Db 658 CACATTTTGGAAAGCCATTATATATCACTGCTCCAAAGCTTATGAAAGCAAGTTCAACAG 717

QY 61 AspLyvAlEulEulEThrgluEulEulEglnHlSseRAlAlYvShlElyvAlAsp11EAsp 80

Db 718 GACAAAGGTAAATTAACAAGACTCATCAAGATTCAAGGAAACATAAGGTGAACATTGAT 777

QY 81 CysSerGIYvAlvAlvAlvAlTYAspQlnSerSerGIvAspVAlAlAserLeuSerSer 100

Db 778 TGCACTCGAAAGGTTGTGATTAGCATCAAGCTCCCAAGATGTGCTCTCTCTTTCA 837

QY 101 AspCyvPheLeuThrVal1EulEulEulYvAlEuglnYvSerPheAsnSerValHisLeu 120

```

Db      838 GACGTTTCTCACTGTAAGTCTGGTAACTGGAGAGAGCTTCAACTGTTTCACTG 897
Qy      121 LeuAAGLYGLYPheAAGLPheSerArgPheProGLYLeuCySGuGLYLYSer 140
Db      898 CTTCAGAGTGGAGTGGTGGTCTCTCGTGTGTTCCCTGGCCCTCTGTGAAGAAATCC 957
Qy      141 ThrLeuValProThrCysIleSerGlnProCysLeuProValAlaAsnIleGLYProThr 160
Db      958 ACCTAGTCCCTACCTGATTTCTCACTTCTGCTTACCTTGTCCAACTGGAGCCACC 1017
Qy      161 ArgIleLeuProAsnLeuTYrLeuGLYCySGlnArgAspValLeuAsnLYSGuLeuIle 180
Db      1018 CGAATTCCTCCCACTTTTATCTTGGCTGCCAGAGATGTCTCAACAAGAGCTGATA 1077
Qy      181 GlnGlnAsnGLYIleGLYTYrValLeuAsnAlaSerTYrThrCysProLYSProAspPhe 200
Db      1078 CAGCAGATGGAGATGGTATGTTAATGATGCAAGCTATACCTGTCCAAAGCTGACTTT 1137
Qy      201 IleProGlnSerHisPheLeuArgValProValAsnAspSerPheCySGuLYSILEu 220
Db      1138 ATCCCGAGTCTCATTTCTCGCTGTGCTGTGATGACAGCTTTTGTGAGAAATTTTG 1197
Qy      221 ProTYrLeuAspLYSLeuValAspPheIleGLYLYSAlaLYSAlaSerAsnGLYCySVal 240
Db      1198 CCGTGGTTGGCAAAATCAGTAGATTTCATGAGAAAGCAAAAGCTCCAAATGAGTGTGT 1257
Qy      241 LeuValHisCysLeuAlaGLYIleSerArgSerAlaThrIleAlaIleAlaTYrIleMet 260
Db      1258 CTAGTGACAGTGTAGTGGAGATCTCCGCTCGCAACATCGCTATCCCTACATCAGT 1317
Qy      261 LysArgMetAspMetSerLeuAspGLYAlaTYrArgPheValLYSGuLYSArgProThr 280
Db      1318 AAGAGATGGACATGATCTTAGATGAAGCTTACAGATTTGTGAAGAAAAAGAACTACT 1377
Qy      281 IleSerProAsnPheAsnPheLeuGLYGLNLeuLeuAspTYrGLYLYSILEYAsn 300
Db      1378 ATATCTCCAAACTTCATATTTCTGGGCCAACTCTGACTATGAGAAAGATTAAAGAC 1437
Qy      301 GlnThrGLYAlaSerGLYProLYSLeuLYSLeuLYSLeuHisLeuHisLYSLeuProAsn 320
Db      1438 CAGACTGGAGCATCAGAGGCCAAAGACCAATCTCACTGTGACCTGGAGAAAGCCAAAT 1497
Qy      321 GlnProValProAlaValSerGLYGLYGLYGLYLYSLeuLYSLeuLYSLeuProSerPro 340
Db      1498 GAACCTGTCCCTGCTCTCTCAGAGGGTGGACAGAAAGAGAGCGCCCTCACTCAACC 1557
Qy      341 CysAlaAspSerAlaThrSerGLYAlaAlaGLYGLNArgProValHisProAlaSerVal 360
Db      1558 TGTGCCGACTCTGCTACCTCAGAGGCGACGAGCAAAAGCCCGTGCATCCCGCAGCGTG 1617
Qy      361 ProSerValProSerValGlnProSerLeuLeuGlnAspSerProLeuValGlnAlaLeu 380
Db      1618 CCCAGGTGCCAGAGCTGACAGCGCTGCTGTTAGAGAGACGCCCGCTGTATACAGGGCTC 1677
Qy      381 SerGLYLeuHisLeuSerAlaAspArgLeuGLYAspSerAsnLYSLeuLYSArgSerPhe 400
Db      1678 AGTGGGCTGCACCTGTCCGACAGAGCTGGAAGACAGCAATAGTCAAGAGTTCTTC 1737
Qy      401 SerLeuAspIleLYSLeuSerValSerTYrSerAlaSerMetAlaAlaSerLeuHisGLY 420
Db      1738 TCTCTGGAATCAAAATCAGTTTCATATTCAGCAGAGATGAGCATCTTACATAGGCTTC 1797
Qy      421 SerSerSerGLYAspAlaLeuGLYTYrTYrLYSProSerThrThrLeuAspGLYThrAsn 440
Db      1798 TCTCTATCAGAAAGATGCTTGGAAATCTCAAACTTCACACTCTGGATGGAGCCAAAC 1857
Qy      441 LysLeuCySGlnPheSerProValGlnLYSLeuSerGLYGLNThrProGlnThrSerPro 460
Db      1858 AAGCTATGCGAGTTCCTCCCTGTTCAAGAACTATCGAGCAACACTCCGAAACAGTCTCT 1917
Qy      461 AspLYSGuGLYAlaSerIleProLYSLeuGLNThrAlaArgProSerAspSerGln 480

```

```

Db      1918 GATTAAGAGAGAACCCAGCATCCCAAGAGCTGCAGACCCGCCAGGCTTTCAGACAGCCAG 1977
Qy      481 SerLYSArgLeuHisSerValArgThrSerSerSerGLYThrAlaGlnArgSerLeu 500
Db      1978 AGCAAGCCATGTCATTCGTCAGAAACAGAGCATGTCACCGCCAGAGGTCCCTTTTAA 2037
Qy      501 SerProLeuHisArgSerGLYSerValGlnAspAsnTYrHisThrSerPheLeuPheGLY 520
Db      2038 TCTCCATGTCATCGAAGTGGAGCGCTGGAGAGCAATTCACACACAGCTTCTTTCCGC 2097
Qy      521 LeuSerThrSerGlnGlnHisLeuThrLYSLeuAlaGLYLeuGLYLeuLYSGLYTYrHis 540
Db      2098 CTTTCACACAGCCAGCAGACACCTCAGAAAGTCTGTGCTGGGCTTAAAGGCTGGCAC 2157
Qy      541 SerAspIleLeuAlaProGlnThrSerThrProSerLeuThrSerSerTYrPheAla 560
Db      2158 TCGATATCTTGGCCCCCAGACCTCTACCCCTTCCGACACAGCAGCTGATTTTCC 2217
Qy      561 ThrGlnSerSerHisPheTYrSerAlaSerAlaIleTYrGLYGLYSerAlaSerTYrSer 580
Db      2218 ACGAGTCTCTACACTTCTACTCTGCTCAGCCATCTACGAGGCGATGCCAGTTACTCT 2277
Qy      581 AlaTYrSerCysSerGlnLeuProThrCysGLYAspGlnValTYrSerValArgArg 600
Db      2278 GCTTACAGCTCAGCAGCAGCTGCCACTTGCAGAGCCAAAGCTATTCTGTGCGCAGCGG 2337
Qy      601 GlnLYSProSerAspArgAlaAspSerArgArgSerTYrPheGLYGLNLeuSerProPheGln 620
Db      2338 CAGAAAGCCAAAGTGCAGAGCTGACTCGCGGCGAGCTGGCATGAGAAAGCCCTTTGAA 2397
Qy      621 LysGlnPheLYSArgArgSerCysGlnMetGlnPheGLYGLNLeuSerIleMetSerGln 640
Db      2398 AAGCATTTAAACCGAGAGAGCTGCCAAATGGAATTTGGAGAGACATCATCTCAGAGAAC 2457
Qy      641 ArgSerArgGLYGLYLeuGLYLYSValGLYSerGlnSerSerPheSerGLYSerMetGln 660
Db      2458 AGGTACCGGGAAGAGCTGGGGAAGTGGGCGAGTCAGTCAGTACTTTTGGCGCAGATGAA 2517
Qy      661 IleIleGlnValSer 665
Db      2518 ATCATTTAGTCTCC 2532

RESULT 3
AY038927      1935 bp      mRNA      linear      PRI 01-AUG-2001
LOCUS        Homo sapiens truncated MAPK phosphatase 7 (MKP7) mRNA, complete
DEFINITION
ACCESSION    AY038927
VERSION      AY038927.1 GI:15072488
KEYWORDS
SOURCE
ORGANISM     Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE    1 (bases 1 to 1935)
AUTHORS      Montpetit,A., Boily,G. and Sinnett,D.
TITLE        A detailed transcriptional map of the chromosome 12p12 tumor
suppressor locus
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 1935)
AUTHORS      Montpetit,A., Boily,G. and Sinnett,D.
TITLE        Direct Submission
JOURNML     Submitted (07-JUN-2001) Hemato-Oncology, Hopital Ste-Justine, 3175
Cote-Ste-Catherine, Montreal, Qc H3T 1C5, Canada
FEATURES
source
1..1935
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
1..1935
/gene="MKP7"
28..459
/gene="MKP7"
CDS

```

```
/note="minor splice variant"  
/codon_start=1  
/product="truncated MAPK phosphatase 7"  
/protein_id="AAK69770.1"  
/db_xref="GI:15072489"  
/translation="MAHEMIQTQIVTERVALLSEGTKEYLLIDSRPVEYNTSHILE  
AIIINCSKLKMRKLODKVLTIELIOMHAKYDIDCSOKRQVYVYDSSQDVASLSDDC  
FLTVLLGKLEKSPNSVHLLAGDAEMDWLCTVCKQGYLSKA"
```

ORIGIN

Alignment Scores:

Pred. No.:	0	Length:	1935
Score:	472.00	Matches:	472
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	70.98%	Indels:	0
DB:	9	Gaps:	0

US-10-029-345A-109 (1-665) x AY038927 (1-1935)

```
QY 194 ThTCyPProlyProApPheHelleProGluSerHisPheLeuArgValProValaAaPp 213  
Db 443 ACCGTGCAAGAGCTGATTTATCCCGAGTCTCATTTCTCGGTGCTGGAATGAC 502  
QY 214 SerPheCyGluLysIleLeuProTyrPLeuApLysSerValaAaPheIleGluLysAla 233  
Db 503 AGCTTTTGAGAAATTTTGGCGGTGGACAAATCAGATTTTCATTGGAAAGCA 562  
QY 234 LysAlaSerAngIlyCyValLeuValHisCyValaAaGlyIleSerArgSerAlaThr 253  
Db 563 AAGCCTCCAAATGATGTGTTAGTGCACCTGTTAGCTGGATCTCCGCTCCGCACAC 622  
QY 254 IleAlaIleAlaIlyrIleLeuLysArgMetAspMetSerLeuAaGluAlaIlyrArgPhe 273  
Db 623 ATGCTATGCTCCATCATCATGAGAGAGATGACATGCTTTAGATAAAGCTTACATTT 682  
QY 274 ValIyGluLysArgProThrIleSerProAngPheAaPheLysGluIleLeuAaPp 293  
Db 683 GTCAAGAAAGAAAGACTTATATCTCCAACTTCAATTTTCTGGCCCACTCTCGAC 742  
QY 294 TyrGluLysLysIleLysAngInThrGlyAlaSerGlyProLysSerLysLeuLysLeu 313  
Db 743 TATGAGAAAGATTAAGAACAGACTGAGCATCGAGCCCAAGAGCAAACTCAAGCTG 802  
QY 314 LeuHisLeuGluLysProAngLysProValProAlaValaSerGluGlyIleLysSer 333  
Db 803 CTGCACTGAGAGAGCCAAATGAACTGCTGCTGCTCAAGAGGTGACGAAAGG 862  
QY 334 GluThrProLysSerProProCyValaAaPheSerAlaThrSerGluAlaIleGlyIlyrArg 353  
Db 863 GAGACGCCCTCTGCTCACTGCTGCGACTTGTCTACCTCAGAGGACGAGCAAAAG 922  
QY 354 ProValHisProAlaSerValProSerValProSerValGlnProSerLeuLeuGluAaPp 373  
Db 923 CCGGTGCATCCGCCAGCGGTGCCAGAGGTGCCAGGCGTGCCTGTTAGAGGAC 982  
QY 374 SerProLeuValGlnAlaLeuSerGlyLeuHisLeuSerAlaAaPpArgLeuGluAaPpSer 393  
Db 983 AGCCCCCTGGTACAGGCGCTCACTGGGCTGCACTGCTCCCAAGACGCTGAGAGCAGC 1042  
QY 394 AaLysLeuLysArgSerPheSerLeuAaPpIleLysSerValaSerIlyrSerAlaAaPp 413  
Db 1043 AATTAAGCTCAAGGCTCTCTCTGATATCAAAATCAATTCATTCAGCACACATG 1102  
QY 414 AlaAlaSerLeuHisGlyPheSerSerSerGluAaPpAlaLeuGluIlyrIlyrLysProSer 433  
Db 1103 GCAGCATCTTACATGCTCTCTCTCATGAGAAAGCTTGGAAATACACAACTTCC 1162  
QY 434 ThrThrLeuAaPpLysArgLysLeuCyGlnPheSerProValGlnGluLeuSerGlu 453  
Db 1163 ACTACTCTGATGGAGCAAAAGCTATGCACTTCTCCCTCTTACAGGAACATATCGGAG 1222  
QY 454 GlnThrProGluThrSerProAaPpLysGluGluAlaSerIleProLysLysLeuGlnThr 473
```

```
Db 1223 CAGACTCCCGAAACCAAGTCTGATTAAGAGAGAGCAGCATCCCAAGAAAGCTGACAC 1282  
QY 474 AlaArgProSerAaPpSerGlnSerIlyrArgLeuHisSerValArgThrSerSerGly 493  
Db 1283 GCCAGGCTTTCAGACAGCCAGAGCAAGCATTTGCTCGTCAAGAACAGCAGCATGGC 1342  
QY 494 ThrAlaArgSerLeuLeuSerProLeuHisArgSerGlySerValGluAaPpAaPp 513  
Db 1343 ACCGCCAGAGGTCCTTTATCTCCACTGCAATCGAAGTGGAGCGTGGAGCAATTC 1402  
QY 514 HisThrSerPheLeuPheGlyLeuSerThrSerGlnGlnHisLeuThrLysSerAlaGly 533  
Db 1403 CACACAGCTTCTCTTTGGCCTTTCACACAGCAGACACCTCAAGAACTGCTGGC 1462  
QY 534 LeuGlyLeuLysGlyTyrPheIleSerAaPpIleAlaProGlnThrSerThrProSerLeu 553  
Db 1463 CTGGGCTTAAAGGCTGCGCACTCGATATCTGGCCCCCAGACCTCTACCTCTCCCTG 1522  
QY 554 ThrSerSerTrpThrPheAlaThrGluSerSerHisPheTyrSerAlaIleTyr 573  
Db 1523 ACCAGAGCTGATATTTTGCACAGATCTTCACACTTCTACTCTGCTCAGCATTCAC 1582  
QY 574 GlyIlySerAlaSerTyrSerAlaTyrSerCySerGlnLeuProThrCyGlyAaPpGln 593  
Db 1583 GAGAGCAGTCCAGATTACTGCTGCTTACAGCTCAGCAGCAGCTGCCACTTGGGAGACAA 1642  
QY 594 ValTyrSerValArgArgArgGluLysProSerAaPpArgAlaAaPpSerArgSerTrp 613  
Db 1643 GTCTATTCTGTGCGCAGCGCGAGAAAGCCAAATGACAGACTGACTCGCGCGAGACTG 1702  
QY 614 HisGluLysSerProPheGluLysGlnPheLysArgArgSerCyGlnMetGluPheGly 633  
Db 1703 CATGAAGAGAGCCCTTTGAAAGCAATTAACGAGAAAGCTGCCAAATGAAATTTGA 1762  
QY 634 GluSerIleMetSerGluAaPpArgSerArgGluGluLeuGlyLysValGlySerGlnSer 653  
Db 1763 GAGAGCATATGTCAGAAACAGGTCAACGGAAGAGCTGGGAAAGTGGGCACTGCT 1822  
QY 654 SerPheSerGlySerMetGluIleIleGluValSer 665  
Db 1823 AGCTTTTCGGCAGCATGGAATCATTTGAGTCTCC 1858
```

RESULT 4

AX260342	1998 bp	DNA	linear	PAT 26-OCT-2001
LOCUS	AX260342			
DEFINITION	Sequence 3 from Patent WO0173059.			
ACCESSION	AX260342			
VERSION	AX260342.1	GI:16509305		
KEYWORDS				
SOURCE				
ORGANISM	Homo sapiens (human)			

REFERENCE

1 Meyers, R.A.
38692 and 21117: dual specificity phosphatase molecules and uses
therefor
Patent: WO 0173059-A 3 04-OCT-2001;
JOURNAL Millennium Pharmaceuticals, Inc. (US)
location/Qualifiers

FEATURES

source	1..1998
	/organism="Homo sapiens"
	/mol_type="unassigned DNA"
	/db_xref="taxon:9606"

ORIGIN

Alignment Scores:	0	Length:	1998
Pred. No.:	472.00	Matches:	663
Score:	99.40%	Conservative:	0
Percent Similarity:	99.40%	Mismatches:	2
Best Local Similarity:	70.98%	Indels:	4
Query Match:			

DB: 6 Gaps: 0
US-10-029-345a-109 (1-665) x AX260342 (1-1998)
QY 1 MetAlHisGluMetIleGlyThrGlnIleValThrGluArgLeuValAlaLeuLeuGlu 20
Db 1 ATGGCCCATGAGATGATTGGAACTCAAAATGTTACTAGAGAGTTGGTGGCTGTGAGAA 60
QY 21 SerGlyThrGluLeuValLeuLeuIleAspSerArgProPheValGluIlyrAntThrSer 40
Db 61 AGTGAACCGAAAGAGCTGTCTAATTGATAGCCGCGCATTTGTGGAATACAAATACATCC 120
QY 41 HisIleLeuGluValAlaIleAsnIleAsnCysSerLeuMetLysArgArgLeuGlnGln 60
Db 121 CACATTTTGGAAACCTTAATATCACTGCTCCAACTTATGAAAGGAAGTTGCAACG 180
QY 61 AspLysValLeuIleThrGluLeuIleGlnHisSerAlaLysHisLysValAspIleAsp 80
Db 181 GACAAAGTGTAAATTACAGAGCTCATCCAGCATTCAGCGAAACATTAAGGTTGACATTTGAT 240
QY 81 CysSerGlnLysValValValTyrAspGlnSerSerGlnAspValAlaSerLeuSerSer 100
Db 241 TGCAGTACAGAGGTGTAGTTACATCAAAAGCTCCCAAGATGTGCTCTCTCTTCA 300
QY 101 AspCysPheLeuThrValLeuLeuGlyLysLeuGluLysSerPheAsnSerValHisLeu 120
Db 301 GACTGTTTCTCATCTGACTTCTGGGTAACCTGGAAGAAGCTTCAACTCTGTTCACCTG 360
QY 121 LeuAlaGlyGlyPheAlaGluPheSerArgCysPheProGlyLeuGlyGlyLysSer 140
Db 361 CTTCGAGGTGGTGTCTGAGTTCCTCGTTGTTTCCCTGGCTCTGTGAAGGAAATCC 420
QY 141 ThrLeuValProThrCysIleSerGlnProCysLeuProValAlaAsnIleGlyProThr 160
Db 421 ACTCTAGTCCCTACCTGCAATTTCTCAGCTTCTGCTTACCTGTGCCAACTGGGCGCAACC 480
QY 161 ArgIleLeuProAsnLeuTyrLeuGlyCysGlnArgAspValLeuAsnLysGluLeuIle 180
Db 481 CGAATTTCTCCCAATCTTATCTGTGGCTCCAGCGAGATGCTCTCAACAGAGAGCTGAT- 539
QY 181 -GlnGlnAsnGlyIleGlyTyrValLeuAsnAlaSerTyr-ThrCysProLysProAsp 200
Db 540 GGAGCGAATGGATGTGTTATGTGTTAAATGCCACAA-TACTCTCCAAAGCCCTGACT 598
QY 200 HeIleProLysSerHisPheLeuArgValProValAsnAspSerPheCysGluLysIleL 220
Db 599 TTATCCCGAGTCTCATTTCTCGCGTGTGCTGTGAATGACAGCTTTTGTGAAGAAATTT 658
QY 220 euProTyrLeuAspLysSerValAspPheIleGluLysAlaLysAlaSerAsnGlyCysV 240
Db 659 TCCCGTGGTTGGCAAAATCAGTAGATTTCATTGAGAAAGCAAAAGCTCCATTCAGATG 718
QY 240 AlLeuValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIleM 260
Db 719 TTTCTATGTGACATGTTAGCTGGGATCTCCGCTCCGCCACCAATCGCTATCGCTACATCA 778
QY 260 eLysArgMetLysMetSerLeuAspGluAlaTyrArgPheValLysGluLysArgProT 280
Db 779 TGAAGAGATGGACATGCTTATGATGAAGCTTACGATTTGGAAAGAAAAAGCCCTA 838
QY 280 hLrIleSerProAsnPheAsnLeuGlyGlnLeuLeuAspTyrGluLysLysIleLysA 300
Db 839 CTATATCTCCAAACTTCATTTTCTGGGCCAACTCCTGGACTATGAGAAAGATTACGA 898
QY 300 snGlnThrGlyAlaSerGlyProLysSerLysLeuLysLeuLeuHisLeuGluLysProA 320
Db 899 ACCAGACTGAGATCAAGGCGCAAGAGCAAACTCAAGCTGTGCACTCGAGAAAGCCA 958
QY 320 snGlnProValProAlaValSerGluGlyGlnLysSerGlnLysProLysSerProP 340
Db 959 ATGAACCTGTCCCTGTGTCTCAAGAGGTGAGACAGAAAGCGAGCCCTCAGTCCAC 1018
QY 340 rCysAlaAspSerAlaThrSerGluAlaAlaGlyGlnArgProValHisProAlaSerV 360

Db 1019 CCTGTGCCACTCTCTACTCTAGAGCGACAGACAAAGGCCCTGCATCCGCGAGGG 1078
QY 360 aLProSerValProSerValGlnProSerLeuLeuGluAspSerProLeuValGlnAlaL 380
Db 1079 TGGCCAGGCGTCCAGCTGCGAGCCGTGCTGTTTGAAGAGACAGCCGCGTGTACAGGCGC 1138
QY 380 euSerGlyLeuHisLeuSerAlaAspArgLeuGluAspSerAsnLysLysLysArgSerP 400
Db 1139 TCAGTGGCTGCACCTGTCCGACAGAGCTGGAAGACAGCAATAGCTCAAGCTCTCT 1198
QY 400 heSerLeuAspLysSerValSerTyrSerAlaSerMetAlaAlaSerLeuHisGlyP 420
Db 1199 TCTCTCGGATATCAATACAGTTTCAATTATTCAGCCAGCATGSCACATCTTACATGGCT 1258
QY 420 heSerSerSerGluAspAlaLeuGluTyrTyrLysProSerThrThrLeuAspGlyThra 440
Db 1259 TCTCTCTCATAGAAATCTTTGGAAATCTTCAAACTTCCACTACTCTGTGATGGAGCA 1318
QY 440 snLysLeuCysGlnPheSerProValGlnGluLeuSerGluGlnThrProGluThrSerP 460
Db 1319 ACAAGCTATGCGAGTTCCTCCCTGTTCAGAACTATCGAGAGACTCCCGAACCAGTC 1378
QY 460 rOAspLysGluGluAlaSerIleProLysLysLeuGlnThrAlaArgProSerAspSerG 480
Db 1379 CTGATTAAGAGAGAACCCAGCATCCCAAGAAAGCTGCACACGCGCAGCTTCACAGACGC 1438
QY 480 LnsrLysArgLeuHisSerValArgThrSerSerSerGlyThrAlaGlnArgSerLeuL 500
Db 1439 AGAGCAACGATGTGATTCGGTCAGAACCAACAGACAGAGGAGACCGCCCAAGAGTCCCTTT 1498
QY 500 euSerProLeuHisArgSerGlySerValGluAspAsnTyrHisThrSerPheLeuPheG 520
Db 1499 TATCTCCACTGCATGAATGGAGACCTGAGAGACAAATTACCAACAGCTTCTTTTCG 1558
QY 520 LysLeuSerThrSerGlnGlnHisLeuThrLysSerAlaGlyLeuGlyLeuLysGlyTyrP 540
Db 1559 GCCTTTCACACAGCAGACAGACCTCAAGAAATGCTGCTGGCTGCTTAAGGCTGGC 1618
QY 540 LnsrSerAspLysLeuAlaProGlnThrSerThrProSerLeuThrSerSerTyrPheA 560
Db 1619 ACTCGATATCTTGGCCCCCGACACTTACCTCCCTTCCACAGCAGCTGTGATTTTTCG 1678
QY 560 LathrGluSerSerHisPheTyrSerAlaSerAlaIleTyrGlyGlySerAlaSerTyrS 580
Db 1679 CCAAGATCTTCACACTTCTACTCTGCTCAAGCATTTACGAGAGCGATGCGCATTTACT 1738
QY 580 eLrAlaTyrSerCysSerGlnLeuProThrCysGlyAspGlnValTyrSerValArgArgA 600
Db 1739 CTGCCTACAGCTGAGCGACGCTGCCCACTTCGAGAGACCAAGTCTATTCTGTGCGCAGGC 1798
QY 600 rGlnLysProSerAspArgAlaAspSerArgArgSerTyrHisGluGluLysProPheG 620
Db 1799 GGCAGAAACCAAGTACAGAACTGACTCGCGCGAGCTGGCAATGMAAGAGCCCTTTTCG 1858
QY 620 LysGlnPheLysArgArgSerCysGlnMetGluPheGlyGluSerIleMetSerGluA 640
Db 1859 AAAACAGTTTAAACGCAAGAGCTGCCAAATGCAATTTGGAGAGCATCTATGTCAGGA 1918
QY 640 snArgSerArgGluGluLeuGlyLysValGlySerGlnSerSerPheSerGlySerMetG 660
Db 1919 ACAAGTCACGGGAAGAGCTGGGGAAGTGGGCAAGTCAAGTCAAGCTTTTTCGGGAGCATAG 1978
QY 660 LulleIleGluValSer 665
Db 1979 AAATCATTTAGAGTCTCC 1995
RESULT 5
AX713989
LOCUS AX713989 2102 bp DNA linear PAT 15-APR-2003
DEFINITION Sequence 673 from Patent EP1293569.
ACCESSION AX713989
VERSION AX713989.1 GI:29888917

QY 580 SerAlaTyrSerCysSerGlnLeuProThrCysGlyAspGlnValTyrSerValArg 599
Db 1793 TCTGCTTACAGCGTGCAGCCAGCGTCCACCTTGCGGAGACCAAGTCATTTCTGTGGCCAGG 1852
QY 600 ATGAGlnlyProSerAspArgAlaAspSerArgArgSerTrpHisGlnGlnSerProPhe 619
Db 1853 CGGCGAAGCCCAAGTACAGAGCTGACTCGCGCGGAGAGCTGGCATTAAGAGAGCCCTTT 1912
QY 620 GlnlyGlnPheIlyAspArgSerCysGlnMetGlnPheGlyGlnSerIleMetSerGln 639
Db 1913 GAAAGACAGTTTAAAGCAGAGAGCTCCAAATGAAATTTTGAGAGAGCATTCATGCAGAG 1972
QY 640 AanaSerArgGlnGlnIleuGlyIlyValGlySerGlnSerSerPheSerGlySerMet 659
Db 1973 AACAGGTCTACGGGAAAGAGCTGGGAAAGTGGGAGTCAGTCTTTCGCGGAGCAGATG 2032
QY 660 GlnIleIleGlnValSer 665
Db 2033 GAAATCATTTGAGGTCTCC 2050
RESULT 6
AK055973
LOCUS
DEFINITION Homo sapiens cDNA FLJ1411 fls. clone NT2NE2000214, moderately
similar to DUAL SPECIFICITY PROTEIN PHOSPHATASE 8 (EC 3.1.3.48).
ACCESSION AK055973
VERSION AK055973.1 GI:16550835
KEYWORDS oligo capping, fls (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
AUTHORS 1 Tashiro,H., Yamazaki,M., Watanabe,K., Kumagai,A., Itakura,S.,
Fukuzumi,Y., Fujimori,Y., Komiyama,M., Sugiyama,T., Irie,R.,
Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,D., Isono,Y.,
Kawai-Hito,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H.,
Matsumoto,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K.,
Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A.,
Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S.,
Nagahara,K., Masuko,Y., Nagai,K. and Isogai,T.
NEDO human cDNA sequencing project
Unpublished
TITLE 2 (bases 1 to 2102)
JOURNAL
REFERENCE
AUTHORS Isogai,T., Otsuki,T. and Sugiyama,T.
DIRECT SUBMISSION
SUBMITTED (24-OCT-2001) Takao Isogai, Helix Research Institute,
Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology (RAB); cDNA library
construction: Helix Research Institute (HRI) (supported by Japan
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
HRI, and Biotechnology Center, National Institute of Technology and
Evaluation; clone selection for full insert sequencing: RAB and
HRI.
FEATURES
source
1. Location/Qualifiers
1..2102
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NT2NE2000214"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/clone_lib="NT2NE2"
/note="cloning vector: pME18SFL3-mRNA from NT2 neuron
after the differentiation of NT2 neuronal precursor
cells."
CDS
56..2053
/note="unnamed protein product"
/codon_start=1

/protein_id="BAB71060.1"
/db_xref="GI:16550835"
/translation="MAHEMIGTQIVTERVALLSEGREKVLIIIDSRPREVYNTSHILE
AINTNCSKLMKRRIQDQVLTITELIQSHAKKVIDCQKVVYVDQSDPDASJSDC
FLTVLGLKERSFNSVHLIAGFAFESFCFLGEGKSLVPTCISQCLPVANIGPT
LILPVLYGECORDVLMKELMOONIGIVLYANASCPKPDFIPESHFLRPVNDSECK
ILPMLDSKVDPIERAKASANGCVLHCLAGISRSATIALAYIMKMDMLDAVRYKE
KRPIISPENFLGGLIDYEKKIKNOTGASGPKLKLHLEKPNPVPVASEGGQKSE
TLGPBCDASATSSAAGORPVHASPVPSPVSDPLVVALSGLIHLSADRLD
SNLKRSPSLDIKVSYSASMAASLIGRSSSEDLLEYREBTTLDTGKLCQFSPVQZ
LSEQTPETSPDKERASIPKLIQTAPSPSQSKRLHVSSTSGTQRLSLPLHNSGS
VEDVHYHSFLGLSTSQOHLTKSAGLGLKMHMSIILAQTSPTSLTSSWYATESHF
YSASAIYGSASAYSQSCPLPTCGDQVYSVRBRQPSDRADSRSRWHEESPFEKQFK
RRSQMERGESIMSENRREELKGVKGSQSPSGSMEIIEVS"

Alignment Scores:
Pred. No.: 0 Length: 2102
Score: 472.00 Matches: 662
Percent Identity: 99.10% Conservative: 0
Best Local Similarity: 99.10% Mismatches: 3
Query Match: 90.96% Indels: 6
DB: Gaps: 0

US-10-029-345A-109 (1-665) x AK055973 (1-2102)

QY 1 MetAlaHisGlnMetIleGlyThrGlnIleValThrGluArgLeuValAlaLeuGln 20
Db 56 ATGGCCCATGAGATGATTGGAATCTAAATTTGATGAGAGGTGTGTGCTGTGGAA 115
QY 21 SerGlyThrGlnIlyValLeuLeuIleAspSerArgProPheValGlnTyrAsnThrSer 40
Db 116 AGTGAAGGAAAGAAAGTCTCTGATTTGATGACCGGCATTTTGGAAATACATCATCC 175
QY 41 HisIleLeuGlnAlaIleAsnIleAsnCySerIlySerIleMetIlyArgArgLeuGln 60
Db 176 CACATTTTGGAAAGCATTAATATCACTCCACAGCTTATGAAAGCGAAGTTGCAACAG 235
QY 61 AspIlyValIleuIleThrGlnLeuIleGlnHisSerAlaIlyHisIlyValAspIleAsp 80
Db 236 GACCAAGGTATTAATACAGAGCTCATCCAGCATTCAGGAAACATTAAGCTTGACATTAT 295
QY 81 CysSerGlnIlyValIlyValIlyAspIleSerSerGlnAspValAlaSerLeuSerSer 100
Db 296 TGCACTCAGAAAGTTGATTTACATCAAGAGCTCCAAAGATGTCCTCTCTCTTCA 355
QY 101 AspCysPheLeuThrValLeuLeuGlnIlyIlySerIlySerPheAsnSerValHisLeu 120
Db 356 GACTGTTTTCACAGTACTTCTGGGTAACTGGGAGAGAGCTTCACTGTTTCACTG 415
QY 121 LeuAlaGlyIlyPheAlaGlnPheSerArgCysPheProGlyLeuCysGlnGlyIlySer 140
Db 416 CTGCAAGGTGGTGTGAGAGTCTCTGTTGTTCCCTGAGCTCTGTGAAGAAATCC 475
QY 141 ThrLeuValProThrCysIleSerGlnProCysLeuProValAlaAsnIleGlyProThr 160
Db 476 ACTCAAGTCCACTGACTGATTTCTGAGCTTGCTTACTGTTGCCAATGGGCAAC 535
QY 161 Arg-IleLeuProAsnLeuIlyLeuGlyCysGlnArgAspValLeuAsnIlyGlnLeu 180
Db 536 CT-AATTTCTCCATCTTTATCTTGGCTGCGAGGAAATGCTCCCAAGAGACTAT 594
QY 180 e-GlnGlnAsnGlyIlyIleGlyTyrValIleuAsnAlaSerTyr-ThrCysProIlyProAsp 199
Db 595 -GAGCAGAAAGGAGTGTATGTGTAAAGCAGAA-TACCTGTCCAAAGCTTAC 652
QY 200 PheIleProGlnSerHisPheLeuArgValProValAsnAspSerPheCysGlnIlyIle 219
Db 653 TTTATCCCGAGATCATTTCTGCTGCTGCTGATGACAGCTTTTGTGAAGAAAT 712
QY 220 LeuProIlyPheAspIlySerValAspPheIleGlnIlyValAlaIlyAsnIlyCys 239
Db 713 TTGCCGTGTGTGACCAATCAGATTTCAATTGAGAAAGCAAAAGCTCCCAATGAGAT 772

QY 240 ValIeuValHisGlyLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIle 259
 DB 773 GTTCTAGTGCATCTTTAGCTGGGATCTCCCGCTCCGACACATCGCTATCGCTCAATC 832
 QY 260 MetIleArgNheAphMetSerIleuAspGluAlaTyrArgPheValIleGlySerIleuMetSerGlu 279
 DB 833 ATGAGAGCATGATGCTTTTAGATGAAGCTTTAGATTTGTGAAGAAAAAGAACCT 892
 QY 280 ThrIleSerProMetPheMetPheLeuGlyGlyIleuLeuAspTyrGlyIleValIleVal 299
 DB 893 ACTATATCTCCAACTTCATTTCTGGGGCAACTCTGACATATGAGAAAGATTTAAG 952
 QY 300 AsnGlyThrGlyAlaSerGlyProIleSerIleuLeuLeuMetIleuGlyIleuValPro 319
 DB 953 AACCAAGCTGGAGCATCAGGCGCAAGCAAACTCAAGCTGTGACCTGGAGAAAGCCA 1012
 QY 320 AsnGluProValProAlaValSerGluGlyGlyIleuLeuSerGluThrProLeuSerPro 339
 DB 1013 AATGAACCTGTCCCTGCTGTCTCAGAGGGTGAACAGAAAGCAGAGCCCTCAAGTCCA 1072
 QY 340 ProGlyAlaAspSerAlaThrSerGluAlaAlaGlyIleuArgProValHisProAlaSer 359
 DB 1073 CCTGTGCCGCACTGTCTACCTCAGAGGCAAGCAAAAGCCCGTGCATCCCGCCAGC 1132
 QY 360 ValProSerValProSerValGlnProSerIleuGluAspSerProLeuValGlnAla 379
 DB 1133 GTTCCCAAGCTGGCCAGCGTGCAGCGCTGTGTAGAGACGCCCGCTGTGACAGCGC 1192
 QY 380 LeuSerGlyLeuHisLeuSerAlaAspArgLeuGluAspSerAsnIleuLeuValAspSer 399
 DB 1193 CTCAGTGGGCTGCACCTGTCTGCAGACAGGCTGGAAGACGCAATTAAGCTCAAGCTTCC 1252
 QY 400 PheSerLeuAspIleuValSerValSerTyrSerAlaSerMetAlaAlaSerIleuHisGly 419
 DB 1253 TTCTCTCTGATATCAATCAATCAATTCATATTCAGCAGCATGCGAGATCTTACATGCG 1312
 QY 420 PheSerSerSerGluAspAlaLeuGluTyrTyrTyrProSerThrThrIleuAspGlyThr 439
 DB 1313 TTCTCTCATCAAGATGCTTTGAAATCTACAACTTCCACTCTCGAGTGGAGACC 1372
 QY 440 AsnIleuLeuGlyGlnPheSerProValGlnIleuSerGluGlnThrProGluThrSer 459
 DB 1373 AACCAAGCTATGCCAGTTCTCCCTGTTCAGGAATCTGAGAGACACTCCGAAACCAAGT 1432
 QY 460 ProAspIleuGluAlaSerIleProIleValLeuGlnThrAlaArgProSerAspSer 479
 DB 1433 CCTGATAGAGGAGAACCGACATCCCAAGAACCTCAGACTCCAGGCTTCAGACAGC 1492
 QY 480 GlnSerIleArgIleuHisSerValArgThrSerSerGlyThrAlaGlnArgSerIleu 499
 DB 1493 CAGAGCAAGGATTTGCTGCTCAAGAACGAGAGAGTGGCACCCCGCAGAGTCCCTT 1552
 QY 500 LeuSerProLeuHisArgSerGlySerValGluAspAsnTyrHisThrSerPheLeuPhe 519
 DB 1553 TTATCTCCACTGATGAGAGTGGAGCGTGGAGACATTAACACACAGCTTCTCTTTC 1612
 QY 520 GlyLeuSerThrSerGlnGlnHisLeuThrIleSerAlaGlyIleuGlyIleuValIleuVal 539
 DB 1613 GGCCTTTCCACAGCAGCAGCAGCACTCAAGAACTGTGCTGGGCTTAAGGAGCTGG 1672
 QY 540 HisSerAspIleuLeuAlaProGlnThrSerThrProSerIleuThrSerSerTyrPhe 559
 DB 1673 CACTCGAATATCTTGGCCCCCAGACCTTCACTCCCTGACGACAGCTGGATTTT 1732
 QY 560 AlaThrGlySerSerHisPheTyrSerAlaSerAlaIleTyrGlyIleSerAlaSerTyr 579
 DB 1733 GCCACAGAGTCTTCAACTTCTGCTGCTGACCATCTACGAGAGGAGTGGCACTTAC 1792
 QY 580 SerAlaTyrSerCysSerGlnLeuProThrCysGlyAspGlnValTyrSerValArgArg 599
 DB 1793 TCTGCTACAGCTGACAGCCAGCTGCCACTTGGGAGAGCAAAAGTATCTGTGGCCAGG 1852

QY 600 ArgGlnIleProSerAspArgAlaAspSerArgArgSerTrpHisGluIleuSerProPhe 619
 DB 1853 CCGCAGAACCCAGATGACAGACTGACTCGCGCGAGAGCTGGCATAAAGAGCCCTTT 1912
 QY 620 GluIleGlnPheIleArgArgSerCysGlnMetGluPheGlyGluSerIleuMetSerGlu 639
 DB 1913 GAAGACAGCTTTAAAGCAGAGAGCTGCCAAATGGAATTTGAGAGAGCATCATGTCAAG 1972
 QY 640 AsnArgSerArgGluIleuGlyIleValIleGlySerGlnSerSerPheSerGlySerMet 659
 DB 1973 AACAGGTACCGGGAAGAGCTGGGGAAGTGGCAGTCAAGTCTTATCTTGGGCGACAGT 2032
 QY 660 GluIleIleGluValSer 665
 DB 2033 GAAATCTTGAAGTCTCC 2050
 RESULT 7
 AX180875 2732 bp DNA linear PAT 06-AUG-2001
 LOCUS Sequence 2 from Patent WO0146394.
 DEFINITION AX180875
 ACCESSION AX180875
 VERSION AX180875.1 GI:15132703
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE
 1 Ploewman, G.D., Martinez, R., Whyte, D., Manning, G., Sudarsanam, S., Hill, R.J., and Flanagan, P. Mammalian protein phosphatases Patent: WO 0146394-A 2 28-JUN-2001; Sugen, Inc. (US)
 FEATURES
 source 1..2732
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 ORIGIN
 Alignment Scores:
 Pred. No.: 0 Length: 2732
 Score: 472.00 Matches: 663
 Percent Similarity: 99.40% Conservative: 0
 Best Local Similarity: 99.40% Mismatches: 2
 Query Match: 70.98% Indels: 4
 DB: 6 Gaps: 0
 US-10-029-345A-109 (1-665) x AX180875 (1-2732)
 QY 1 MetAlaHisGluMetIleGlyThrGlnIleValIleThrGluArgLeuValAlaLeuLeuGlu 20
 DB 538 ATGGCCCATGAGATGATTTGAACTCAAAATTTGACTGAGAGCGTGTGCTGTGCGGAA 597
 QY 21 SerGlyThrGlnIleValLeuLeuIleAspSerArgProPheValGluIleThrSerArg 40
 DB 598 AGTGAAGGAAAAGTGTCTTAATGTATGACCGGCACTTGTGGAATACAAATACATCC 657
 QY 41 HisIleLeuGluAlaIleAsnIleAsnCysSerIleuMetIleuValArgArgLeuGlnGln 60
 DB 658 CACATTTTGAAGCCATTATATCAATGCTCCCAAGCTTATGAAGGAAAGTTGCAACAG 717
 QY 61 AspIleValLeuIleThrGluLeuIleGlnHisSerAlaIleValHisValAspIleAsp 80
 DB 718 GACAAAGTGTAAATTAACAGACTCATCCAGATTCAGCGAAACATTAAGGTTGACATTGAT 777
 QY 81 CysSerGlnIleValIleValIleThrArgPheIleuSerGlnAspValAlaSerLeuSerSer 100
 DB 778 TGCAGTCAAGAGTGTGATTTACATCAAAAGCTCCCAAGAGTGGCTCTCTCTTCA 837
 QY 101 AspCysPheLeuThrValLeuLeuGlyIleValLeuGluIleValSerPheAsnSerValHisLeu 120
 DB 838 GACTGTTTCTCACTGTACTCTGGGTAAACTGGAGAAAGACTTCACTCTTCACTG 897

QY	121	LeuAlaGlyGlyPheAlaGluPheSerArgCysPheProGlyLeuGlyLysSer	140
DB	898	CTTCAGAGTGGATTGCTGAGTTCTCTGTTGTTCCCTGGCCCTGTGAGGAAATTC	957
QY	141	ThrLeuValProThrCysAlaSerGlnProCysLeuProValAlaSerIleGlyProThr	160
DB	958	ACTCTAGTCCCTACTGATTTCTCAGCTTGTCTTACTGTGGCAACTTGGGCCAAC	1017
QY	161	ArgIleLeuProAlaLeuIleuTyrLeuGlyCysGlnArgAspValLeuAsnLysGluLeuIle	180
DB	1018	CGAATTTCTCCCAATCTTTATCTTGGCTGCAGGAGATGTCTCAACAAGGAGCTGAT	1076
QY	181	-GlnGlnAsnGlyIleGlyTyrValLeuAsnAlaSerTyr-ThrCysProLysProAsp	200
DB	1077	GCAGCAGATGGAGATTGGTTATGTGTTAAATGCCAGCA-TACCTGTCCAAAGCCTGACT	1135
QY	200	heIleProGluSerHisPheLeuArgValProValAsnAspSerPheCysGluLysIleI	220
DB	1136	TTATCCCGAGTCTCATTTCTGCTGCTGCTGATGATGACACTTTTGTGAGAAATTT	1195
QY	220	eulProLysLeuAspLysSerValAspPheIleGlyLysAlaLysAlaSerAsnGlyCysV	240
DB	1196	TGCGGTGTGGACAAATCAGTATTTCAATTGAGAAAGCAAAAGCCTCAATGAGATGTG	1255
QY	240	alleuValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIleM	260
DB	1256	TTCTAGTGCATGTTTATAGTGGATCTCCGCTCCGCCACATGCTTATGCTTACATCA	1315
QY	260	eLysArgMetAspMetSerLeuAspGluAlaTyrArgPheValLysGluLysArgProT	280
DB	1316	TGAAGGAGATGACATGCTTTAGATGAGCTTACAGATTGTGAAAGAAAAAGACTTA	1375
QY	280	hrlIleSerProAsnPheAsnPheLeuGlyGlnLeuLeuAspTyrGlyLysIleLysA	300
DB	1376	CTATATCTCCAACTTCAATTTCTG3GCCCACTCTCGACATATGAGAAAGATTMAA	1435
QY	300	snGlnThrGlyAlaSerGlyProLysSerLysLeuLysLeuHisLeuGluLysProA	320
DB	1436	ACCAAGCTGGAGATCAGGGCCCAAGAGCAAACTCAAGCTGTGACCTGGAGAGACCA	1495
QY	320	snGluProValProAlaValSerGlyGlyGlnLysSerGlnThrProLeuSerProP	340
DB	1496	ATGAACTGTCTGCTGTCTCAGAGGCTGACAGAAAGCAGAGAGCCCTCAGTCCAC	1555
QY	340	roCysAlaAspSerAlaThrSerGluAlaAlaGlyLysIleProValHisProAlaSerV	360
DB	1556	CCGTGCCCACTGTGCTACTCAGAGGACAGAGCAAAAGCCCGCTGCAATCCCGCAGCG	1615
QY	360	alProSerValProSerValGlnProSerLeuLeuGlnAspSerProLeuValGlnAlaL	380
DB	1616	TGCCACGCGGCCACAGCGTGCAGCGCTGTTAGAGACAGCCCGCTGTATCACGGCC	1675
QY	380	eulSerGlyLeuHisLeuSerAlaAspArgLeuGlnAspSerAsnLysLeuLysArgSerP	400
DB	1676	TGAGTGGGTGACCTGTGCGGACAGAGCTG3AAGACAGCATTAAGCTCAAGCGTTCT	1735
QY	400	heSerLeuAspIleLysSerValSerTyrSerAlaSerMetAlaLysLeuHisGlyP	420
DB	1736	TCTCTGTGATATCAATATGATTTTCAATATTCAGCAGCAATGGCAGCATCTTACATGCT	1795
QY	420	heSerSerSerGlnAspAlaLeuGluTyrTyrLysProSerThrThrLeuAspGlyTyrA	440
DB	1796	TCTCCTCATCAGAAAGTGTCTTGGAAATCAAAACCTTCCACTCTGATGGAGACCA	1855
QY	440	snLysLeuCysGlnPheSerProValGlnGlnLeuSerGlnGlnThrProGlnThrSerP	460
DB	1856	ACAACCTAATGCAAGTCTCCCTGTTGAGAACTATCGAGACGACTCCGCAACCAATGC	1915
QY	460	roAspLysGlnGlnLysAlaSerIleProLysLysLeuGlnThrAlaArgProSerLysP	480
DB	1916	CTGATTAAGAGGAAAGCCAGCATCCCAAGAGAGCTCAAGATGCCAGGCTTTCAGACGCC	1975

QY	480	InSerLysArgLeuHisSerValArgThrSerSerSerGlyThrAlaGlnArgSerLeuL	500
DB	1976	AGAGCAAGGATGTGATTCGTCAGAAACAGCAGAGAGGCAACCCCAAGAGTCCCTTT	2035
QY	500	eulSerProLeuHisArgSerGlySerValGlnAspAsnTyrHisThrSerPheLeuPheG	520
DB	2036	TATCTCCATGATCGAAGTGGAGCGTGGAGAGCAATTCACACCAAGCTTCTTTTCG	2095
QY	520	LysSerThrSerGlnGlnHisLeuThrLysSerAlaGlyLeuGlyLysLysGlyTyrP	540
DB	2096	GCTTTCACACGACGACAGACCTCAAGAGTGTGCTGCGCTTAAAGGCTGCG	2155
QY	540	LeSerAspIleLeuAlaProGlnThrSerThrProSerLeuThrSerSerThrTyrPheA	560
DB	2156	ACTCGAATATCTGACCCCAAGACTTACCTCCCTCAACCAAGCGTGAATTTTG	2215
QY	560	laThrGlnSerSerHisPheTyrSerAlaSerAlaIleTyrGlyGlySerAlaSerTyrS	580
DB	2216	CCACAGATCTCTCAACATTTACTGTGCTCAGCCTTACGAGAGCGAGTCCAGTTACT	2275
QY	580	eralTyrSerCysSerGlnLeuProThrCysGlyAspGlnValTyrSerValArgArgA	600
DB	2276	CTGCTTACAGCTGCAGCAGCTGCCACTTGCAGACCAAGTCTATTCTGTGCGAGGC	2335
QY	600	rgGlnLysProSerAspArgAlaAspSerArgArgSerThrHisGlnGluSerProPheG	620
DB	2336	GGCAGAACCAAGTACAGACTACTGCGCGAGCTGGCAGAGAAAGAGCCCTTTG	2395
QY	620	LysGlnPheLysArgArgSerCysGlnMetGlnPheGlyGluSerIleMetSerGluA	640
DB	2396	AAAGCAATTTAAACGAGAAAGCTGCCAAATGGAATTTGGAGAGATCATGTCAAGA	2455
QY	640	snArgSerArgGlnGluLeuGlyLysValGlySerGlnSerSerPheSerGlySerMetG	660
DB	2456	ACAGGTCAACGGAAGAGCTGGGGAAGTGGCACTCATGCTTTTCCGGGACATAGG	2515
QY	660	luIleIleGluValSer	665
DB	2516	AAATCATTTAGAGTCTCC	2532

RESULT 8	AX278461	3059 bp	DNA	linear	PAT 02-NOV-2001
LOCUS	AX278461				
DEFINITION	Sequence 1 from Patent WO0177340.				
ACCESSION	AX278461				
VERSION	AX278461.1	GI:16605915			
KEYWORDS					
SOURCE					
ORGANISM	Homo sapiens (human)				
REFERENCE	1	Duecker, K.			
AUTHORS		Identification of a dual specificity phosphatase: dusp-10			
TITLE		Patent: WO 0177340-A 1 18-Oct-2001;			
JOURNAL		MERCK PATENT GmbH (DE)			
FEATURES					
source		Location/Qualifiers			
CDS		1..3059			
		/organism="Homo sapiens"			
		/mol_type="unassigned DNA"			
		/db_xref="taxon:9606"			
		127..2124			
		/note="unamed protein product"			
		/codon_start=1			
		/protein_id="CAD10545.1"			
		/db_xref="GI:16605916"			
		/db_xref="REMBTREMBL:CAD10545"			
		/translation="MAHEMIGTQIVTRVLALLESSTKXVLIDSRPFVNTSHILE			
		AININSGKLMKRLQODKVLITBLIQHSAGHKVDIOSQVAVVYDSSODVSLSDC			
		FLTVLLGCKEKSFPNSVHLAAGFAEFSRCPGCEKSKTLVPTCLSQPLCPVANIPT			
		RIIPNLVYLGORDVNLKRELMOQNGIYVNASNTCPKPFIPESHRLRPVNDSPCEK			
		ILPWLDRSVDPIEKASNSGCVLVHCLAGISRSATIAIYVIMKRMDSIDEAYRFSKE			
		KRPITSPNFVFLQGLLDYEKKIKNGTGAAGSPSKLTLHLKRNRPVPAVSGQKSE			

ORIGIN

TPUSPPCADSAISSEAQRVHPHVASVPSVQSLLEDSPLVQALSGHLGADRIED
 SNLKRSFSLDIKSVSYASMASLHGFSSBEDALRYKPTSLDNTNLCOFSPQOS
 LSEOTETSPDKESAI PKKLCIAPASDSKSLHSRISSTADRSALSLPSHREGS
 VEDNHTSPLFGISTQOHLTSAGLGYGMSHDILAPOTSPRLTSSVPTSSHF
 YASASITGGSASIASCSQULPCGQVYSVRROKPSRADSRKSHSESPKQFK
 RNSQMEFBSIWSNRSRELEKGVSSFSGSMEIIVS"

Alignment Scores:

Pred. No.:	0	Length:	3059
Score:	472.00	Matches:	663
Percent Similarity:	99.40%	Conservative:	0
Best Local Similarity:	99.40%	Mismatches:	2
Query Match:	70.98%	Indels:	4
DB:	6	Gaps:	0

US-10-029-345a-109 (1-665) x AX278461 (1-3059)

QY 1 MetAlHiegluMetIleGlyThrgInIleValThrgInuArgLeuValAlaleuLeuGlu 20
 DB 127 ATGGCCCATGAGATGATGGACTCAATGTTACTAGAGAGTGGTGCTGCTGGAA 186
 QY 21 SerGlyThrgInuValLeuLeuIleApsSerArgProPheValGluTyraThrSer 40
 DB 187 AGTGAACGGAAAAAGTGTCTAATGATAGCCGGCCATTGTGGAATCAATACATCC 246
 QY 41 HieIleleuGluAlaIleAenIleAenCysSerIleuMetIleYargArgLeuGlnGln 60
 DB 247 CACATTTTGGAAACCATTAATATCAACTGCTCCAACTTAAGAAGGAAGTTGCAACAG 306
 QY 61 AspIleValLeuIleThrgInuLeuIleGlnHieSerAlaYshIshIshValAapIleAap 80
 DB 307 GACAAAGTGTATTTACAGAGCTCATCCAGCATTCAGCGAAACATTAAGTGTGACATTGAT 366
 QY 81 CysSerGlnuValValValTyraApgInSerSerGlnAapValAaserIleuSerSer 100
 DB 367 TGCAGTCAGAAAGGTTGATTAAGATCAAAAGCTCCCAAGATGTGCTCTCTCTCA 426
 QY 101 AspCysPheLeuThrValLeuLeuGlyYshLeuGlnuIlySerPheAenSerValHieIleu 120
 DB 427 GACTGTTTTCTCACTGACTCTGGGTAACCTGGAAGAGCTTCAACTGTTCACCTG 486
 QY 121 LeuAlaGlyGlyPheAlaGluPheSerArgCysPheProGlyLeuGlyGlyIlySer 140
 DB 487 CTTCGACGGTGGTTGCTGAGTCTCTCTGTTTCCCTGCTGCTGTAAGAAATCC 546
 QY 141 ThrLeuValProThrCysIleSerGlnProCysLeuProValAlaAenIleGlyProThr 160
 DB 547 ACTCTAGTCCCTACCTGCAATTTCTCAGCCTTGCTTACTGTTGCCAAACATTTGGGCCAAC 606
 QY 161 ArgIleLeuProAenLeuIlyLeuGlyCysGlnAargAapValLeuAenIlyGluLeuIle 180
 DB 607 CGAATTTCTCCCAATCTTTATCTTGGCTGCCAGCAGATGTCTCAACAAGAGAGCTGAT- 665
 QY 181 -GlnGlnAenGlyIleGlyTyraValLeuAenAlaSerTyraThrCysProIlyProAap 200
 DB 666 GCGACGAAGAGGATGGTATGTATGTAAATGCCAGAA-TACCTGTCCAAACCTCGACT 724
 QY 200 heIleProGluSerHiePheLeuArgValProValAenApsSerPheCysGlyIlyIleIle 220
 DB 725 TTATCCCGGAGTCTCATTTCTCGGTGCTGCTGATGACACCTTTTGTGAGAAAATTT 784
 QY 220 eubProIleuAapIlySerValAapPheIleGlyIlyValAlaYshAserAenGlyCysY 240
 DB 785 TGGCGGTGTGGAACAATCAGTATTCATTGAGAAAGCAAAAGCTCCCAATGAGATGIG 844
 QY 240 alLeuValHieCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyraIle 260
 DB 845 TTCTAGGACCTGTATTAAGTGGGATCTCCGCTCCGACCATCGCATCGCTTACATCA 904
 QY 260 etIlyAargMeCAspMetSerLeuAapGluAlaTyraArgPheValIlyGlyIlyAargPro 280
 DB 905 TGAAGAGATGACATGTCTTTAGATGAACCTTACAGATTTGTGAAGAAAAGAAAGACCTA 964

QY 280 hrIleSerProAenPheAenPheLeuGlyGlnLeuLeuAapTyraGlyIlyIleIlyA 300
 DB 965 CATATCTCCAAACTTCAATTTTCTGGGCCAACTCCTGACATATGAAAGAAATTAAGA 1024
 QY 300 enGlnThrGlyAlaSerGlyProIlySerIlyLeuIlyIleLeuIlyIleLeuIlyPro 320
 DB 1025 ACCAGACTGAGACATCAGCGCCAAAGAGCAAACTCAAGCTGCTGACCTGAGAACCCAA 1084
 QY 320 enGluProAlaProAlaValAserGluGlyGlnIlySerGluThrProIlySerPro 340
 DB 1085 ATGAACCTGTCTCTGTGTCTTGAAGGGGTGACAGAAAAGCAAGAGCCCTCAGTCCAC 1144
 QY 340 roCysAlaApsSerAlaThrSerGluAlaAlaGlyGlnArgProValHieProAlaSer 360
 DB 1145 CCTGTCCGACCTCTGTACTCTAGAGGCGACGACGACAAAGCCCGTGCATCCCGCACGG 1204
 QY 360 alProSerValProSerValGlnProSerLeuLeuIlyApsSerProLeuValGlnAla 380
 DB 1205 TGGCCAGCGTCCAGCGTGCAGCGTCCGCTGTAGAGAGACCGCGGTATACAGGCGC 1264
 QY 380 eusSerGlyLeuHieIleuSerAlaAapArgLeuGluApsSerIlyLeuIlyIlyAargSer 400
 DB 1265 TCAGTGGCTGACCTGTCCGAGACAGGCTGGAAGACGAAATAGCTCAAGCCTTCT 1324
 QY 400 heSerIleuAapIleIlySerValSerTyraSerAlaSerMetAlaAaserIleuHieGly 420
 DB 1325 TCTCTTGATATTCANAATCATGTTTCAATTCAGCCAGCATGGCAGATCTTACATGCT 1384
 QY 420 heSerSerSerGlnuAapAlaLeuGluTyraIlyAapSerThrThrLeuAapGlyThra 440
 DB 1385 TCTCTCATCAGAGATGCTTTGGAATACAAACCTTCACTACTCTGATGGAGCA 1444
 QY 440 enIlyLeuGlyGlnPheSerProValGlnGluLeuSerGluGlnThrProGluThrSer 460
 DB 1445 ACAAGCTATGCAAGTCTCCCTGTTACAGAACTATCGAGAGACATCCCGAAACAGATC 1504
 QY 460 roAapIlyGluGluAlaSerIleProIlyIlyLeuGlnThrAlaAargProSerAapSer 480
 DB 1505 CTGATAGAGAGAGAGCAGACATCCCAAGAGCTGACAGCCGACGCTTCAGACAGCC 1564
 QY 480 InSerIlyAargLeuHieSerValArgThrSerSerSerGlyThrAlaGlnAargSerLeu 500
 DB 1565 AGAGCAGAGGATTCATTCGTGTGACAGACAGACAGATGGACCGCCAGAGGTCCCTTT 1624
 QY 500 eusSerProLeuHieAargSerGlySerValGluAapAenTyraIshThrSerPheLeu 520
 DB 1625 TATCTTCATCGATCGAAGTGGAGCGTGGAGGACAAATTAACACACACACTTCTTTG 1684
 QY 520 IlyLeuSerThrSerGlnGlnHieIleuThrIlySerAlaGlyLeuGlyIlyLeuIlyIly 540
 DB 1685 GCCTTTCCACGACGACGACGACCTTCACGAATCTGTGCTGCTGCTTAAAGGCTGAC 1744
 QY 540 IserAapIleLeuAlaProGlnThrSerThrProSerLeuThrSerSerTyraPhe 560
 DB 1745 ACTCGATATCTTGGCCCCCAGACTCTTACCTTCCCTGACAGCAGCTGGTATTTTG 1804
 QY 560 IaThrGluSerSerHiePheTyraSerAlaSerAlaIleTyraGlyIlySerAlaSerTyra 580
 DB 1805 CCACAGAGCTTCACACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1864
 QY 580 eAlaTyraSerCysSerGlnLeuProThrCysGlyAapGlnAlaTyraSerValAargArg 600
 DB 1865 CTGCTTACAGCTGACGACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1924
 QY 600 rgGlnIlyProSerAapArgAlaApsSerArgArgSerTrpHieGluGluSerProPhe 620
 DB 1925 GGCAGAGCCAGAGTACAGAGCTGACTCCGCGGAGCTGCTGCTGCTGCTGCTGCTGCT 1984
 QY 620 IlyLeuGlnPheIlyAargArgSerCysGlnMetGluPheGlyGluSerIlyMetSerGlu 640
 DB 1985 AAAAGCAGTTTAAAGCAGAGAGCTGCAAAATGGAATTTGGAAGAGAGCATCATGTGAGA 2044

QY 640 snArGerArGluGluLeuGlyLeValGlySerGlnSerSerPheSerGlySerMetC 660
|
|
|
Db 2045 ACAAGGACAGGAGAGAGCTGGGAAAGTGGGACGACGCTAGCTTTTGGGACAGATGG 2104
|
|
|
QY 660 luileleGluValSer 665
|
|
|
Db 2105 AAATCATGTAGGTCTCC 2121
|
|
|
RESULT 9
AX405700
LOCUS AX405700 3104 bp DNA linear PAT 14-JUN-2002
DEFINITION Sequence 115 from Patent WO0222660.
ACCESSION AX405700
VERSION AX405700.1 GI:21438839
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Tang, Y.T., Liu, C., Zhou, P., Asundi, V., Zhang, J., Zhao, Q.A., Ren, F.,
Xue, A.D., Yang, Y., Wehrman, T. and Drmanac, R.T.
TITLE Novel nucleic acids and polypeptides
JOURNAL Patent: WO 0222660-A 115 21-MAR-2002;
HSEQ, INC. (US)
FEATURES
source Location/Qualifiers
1..3104
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
159..2156
/note="unnamed protein product"
/codon_start=1
/protein_id="CAD34807.1"
/db_xref="GI:21438840"
/db_xref="EMBL:CAD34807"
/translation="MAHEMIGTQIVTERLVALLSEGTKEVLLIDSRPFVEYNTSHILE
ANINCSKMKRLQODKYLITELIOHSAKHKVIDSCQVYVVOSSDVASLSDC
FLTVLLGKLEKSPFNHVLLAGAPSRFPGLCEGSKTLVPTCI SOPCLPANIAGPT
RLPNLYLACGQDVINKELMQONGICGYVNAANTCPKPFITESHFLRPVNDSCCK
LLPMDKSYDFLEKAKASNGCVLVHCLAIKRSATTAIAVIMKRMDSIDEXARVKE
KRPTISPENFLGQLDYERKIKNOGAGAPSKKLLLEKPNRPVPAVSGGQKSE
TPLSPPCADASTSEAGRPVHPASVPSVOPSLIEDSPVQALSGIHSADILED
SNKRSFSLDIKSYVSASMAASHGFSSSDALEYKPSFTLLDGTNLLCOFSVOE
LSQRTPEPSPKDEPASI PKKIQOTARPSPDSKRIHSVRSSTGTORSLISLHSGS
VEDNYHSTRLEGLSTSQOHLUTSAGIGLKGMSDILAPOTSPSITSMYRTESHSE
YASAIYGGASYSAYSCQPLPTCDQVTVRRKQPPSDRADSRKSRMSWHEESPFEKQFK
RRSCOMEFGEISIMENSRRELGKVSQSSFGSMEILIEVS"

ALIGNMENT SCORES:
Pred. No.: 0 Length: 3104
Score: 472.00 Matches: 663
Percent Similarity: 99.40% Conservative: 0
Best Local Similarity: 99.40% Mismatches: 2
Query Match: 70.98% Indels: 4
DB: 6 Gaps: 0

US-10-029-345A-109 (1-665) x AX405700 (1-3104)

QY 1 MetAlaHleGluMetIleGlyThrGlnIleValThrGluArgLeuValAlaLeuLeuGlu 20
|
|
|
Db 159 ATGGCCCAATGAGATGATGGAATCTCAATTTGTTACTGAGAGGTGGTGGCTCTGCTGGAA 218
|
|
|
QY 21 SerGlyThrGluLeuValLeuLeuIleAAsPSeArGpProPheValGluTyrAnThrSer 40
|
|
|
Db 219 AGTGGAAACGGAAGAAAGTGTGCTCAATTGATGAGCCGCGCAATTTGTGGAATACATATCATCC 278
|
|
|
QY 41 HisIleLeuGluAlaIleAsnIleAsnGlySerSerLeuMetLeuYsArgArgLeuGlnGln 60
|
|
|
Db 279 CACATTTTGGAAAGCCATTATATCAACTCTCCAGCTTATGAGGAAAGGTGGCAACAG 338
|
|
|
QY 61 AspLysValIleuIleThrGluLeuIleGlnHisSerAlaLeuHisIleValAspIleAAsp 80
|
|
|

Db 339 GACAAAGGTATATATACAGAGCTCATCCAGATTACAGGAAACATTAAGTTGACATTAAT 398
|
|
|
QY 81 CysSerGlnLysValIleValIleTyrAspGlnSerSerGlnAspValAlaSerLeuSer 100
|
|
|
Db 399 TGCAGTTCAGAAAGTGTGATGTTACGATCAAGAGCTCCCAAGATGTGCTCTCTCTTCA 458
|
|
|
QY 101 AspCysPheLeuThrValIleuLeuGluLysLeuGluLysSerPheAsnSerValHisLeu 120
|
|
|
Db 459 GACTGTTTCTCAGCTGACTCTTCTGGGTAACTCGAAGAAAGAGTTCACTCTGTTACCTG 518
|
|
|
QY 121 LeuAlaGlyGlyPheAlaGluPheSerArgCysPheProGlyLeuCysGluGlyLysSer 140
|
|
|
Db 519 CTTCAGAGTGGGTTTGTCTGAGTGTCTCTGTTGTTCCCTGGGCTCTGGAAGAAATCC 578
|
|
|
QY 141 ThrLeuValProThrCysIleSerGlnProCysLeuProValAlaAsnIleGlyProThr 160
|
|
|
Db 579 ACTCTAGTCCCTACCTGATTTCTCAGCCTTGCTTAACCTGTTGCCAACATTTGGGCAC 638
|
|
|
QY 161 ArgIleLeuProAsnLeuTyrLeuGluLysGlnArgAspValLeuAsnLysGluLeuIle 180
|
|
|
Db 639 GCAATTTCTCCCAATCTTATCTTGGCTGCCAGCAGATGTCCTCAACAGAGCTGAT- 697
|
|
|
QY 181 -GlnGlnAsnGlyIleGlyTyrValIleuAsnAlaSerTyr-ThrCysProLysProAAsp 200
|
|
|
Db 698 GCAGCAGATGGAGATTGGTTATGTGTAAATGCCAGAA-TACCTGTCCAAAGCTGACT 756
|
|
|
QY 200 heileProGluSerHisPheLeuArgValProValAsnAspSerPheCysGluLysIleLe 220
|
|
|
Db 757 TTAATCCCGAGCTCATATTTCTGCGCTGTGCTGTGAAGACAGCTTTGTGAAAAATTT 816
|
|
|
QY 220 euProTyrLeuAspLysSerValAAspPheIleGluLysAlaLysAlaSerAsnGlyCysV 240
|
|
|
Db 817 TCCCGTGTGGACAAATCAGTAGATTTCATTGAGAAAGCAAAAGCTTCCATGAGATGTG 876
|
|
|
QY 240 aIleuValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIleM 260
|
|
|
Db 877 TTCTAGTCACATGTTTACTGGAGATCTCCCGCTCGCACCATCGCTATCCGCTACATCA 936
|
|
|
QY 260 eLysArgMetAspMetSerLeuAspGluAlaTyrArgPheValLysGluLysArgProT 280
|
|
|
Db 937 TGAAGAGATGACATGCTTATGATGAGCTTACAGATTTGTGAAAGAAAAAACCTTA 996
|
|
|
QY 280 hrIleSerProAsnPheAsnPheLeuGlyGlnLeuLeuAspTyrGluLysLysIleLysA 300
|
|
|
Db 997 CTATATCTCCCACTTCAATTTTCTGGGCCAATCTTGACATAGAAAGAAATTAAGA 1056
|
|
|
QY 300 snGlnThrGlyAlaSerGlyProLysSerLysLeuLysLeuLeuHisLeuGluLysProA 320
|
|
|
Db 1057 ACCAGACTGGAGCATCAGGCGCAAGAGCAAACTCAAGCTGCTGCACTGGAGAAAGCCAA 1116
|
|
|
QY 320 snGluProValProAlaValSerGluGlyGlyGlnLysSerGluThrProLeuSerProP 340
|
|
|
Db 1117 ATGAACCTGTGCTCTGCTCTCTCAGAGGGTGAACAAAAGCAGAGCGCCCTCACTTCCAC 1176
|
|
|
QY 340 roCysAlaAspSerAlaThrSerGluAlaAlaGlyGlnArgProValHisProAlaSerV 360
|
|
|
Db 1177 CCTGTGCCGACTCTGCTACTCCTCAGAGCAGCAGCAAGAGCCCGCTGATCCCGCAGCG 1236
|
|
|
QY 360 aIProSerValProSerValGlnProSerLeuLeuGluAspSerProLeuValGlnAla 380
|
|
|
Db 1237 TGCCAGAGGTGCCAGCTGACAGCCGCTGTTAGAGAGACAGCCCGCTGATACAGCGC 1296
|
|
|
QY 380 euSerGlyLeuHisLeuSerAlaAspArgLeuGluAspSerAsnLysLeuLysArgSerp 400
|
|
|
Db 1297 TCAGTGGCTGTGACCTGTGCCAGCAGCTGGAAGACAGCAATTAAGCTCAAGCGTTCT 1356
|
|
|
QY 400 heSerLeuAspIleLysSerValSerTyrSerAlaSerMetAlaIleSerLeuHisGlyP 420
|
|
|
Db 1357 TCTCTCTGATATCAATCACTTCAATTTCAAGCAGCATGGCAGCATCTTACATGGCT 1416
|
|
|
QY 420 heSerSerSerGluAspAlaLeuGluTyrTyrLysProSerThrThrLeuAspGlyThra 440
|
|
|

Db 1417 TCTCTCATCAGAGATGCTTTGGAACTACTACAACTCTCACTGATGGAGCA 1476
Qy 440 ANLYLEUCYSGINPHESEPRoVALGInLUeUSerGIuGINThrPROGInLUThSerP 460
Db 1477 ACAAGCTATGCGAGTCTCCCTGTTGAGGAATATCGAGAGACTCCGAAACCAAGTC 1536
Qy 460 roApLySGInLUaLaseRILeProLybLSeuGINThrLAaRgProSeRAspSeRg 480
Db 1537 CTGATTAAGAGAGAACCGAGATCCCAAGAACTGCAAGCCGAGGCTTCAGAAAGCC 1596
Qy 480 InSeRlybARGLeUHiSeRValARgThrSeSeSeRgIYThrLAaInARSeRLeU 500
Db 1597 AGAGCAAGCGATTCATTCGGTCAGAAACGAGACAGTGGACACGCCCAAGAGTCCCTTT 1656
Qy 500 eUSerProLeuHiSaRgSeRGIySeRValGIuApHeNtYrHISThrSeRPhLeuPhg 520
Db 1657 TATCTCCACAGCTCAAGAGGAGCGTGAGAGAAATTAACACACAGCTTCCTTTTGG 1716
Qy 520 lYLeuSeRThrSeRGIInHISLeUThrLySeRValaGIyLeuGIYLeuYGIYTrpH 540
Db 1717 GCCTTCCACACCGAGCAAGACCTCAAGAGTCTGCTGGCTGGGCTTTAGGGCTGGC 1776
Qy 540 iSeRAspILeUaLAPROGInThrSeRThrProSeRLeUThrSeRSeRTrpYrPha 560
Db 1777 ACTGGATATCTTGCCCCCAAGACCTCAACCTTCCTGACAGAGAGTGGTATTTTG 1836
Qy 560 lAthrGLuSeRSeRThrPhetYrSeRAlaSeRAlaIeTYrGIyGIYSeRAlaSeRTYs 580
Db 1837 CCAACAAGTCTCAACCTTCTACTGCTGCTCAACCACTCAAGAGGAGGAGCCAGTTACT 1896
Qy 580 eRAlATYrSeRSeRSeRGIInLeUProThrCySeGIYAPGInVALTYrSeRValARgAR 600
Db 1897 CTGCTTACAGCTCAGCAAGCTCCCACTTGGAGAGCAAGTCTATCTGTGCGAGG 1956
Qy 600 rGInLybProSeRAspARgAlaAspSeRARgARSeRTrpHISGIuGIuSeRProPhg 620
Db 1957 GCGAGAGCCAACTGACAGAGCTGACTCGCGCGAGCTGGCATGAAGAGCCCTTTG 2016
Qy 620 lULySeGInPhelybARgARSeRSeRGIInMeTGLuPhEGIYUSeRlIeMeSeRGIuA 640
Db 2017 AAAAGCAATTAAACCAAGAGCTGCCAAATGCAATTTGAGAGAGCATCATCTCAGAG 2076
Qy 640 sARgSeRARgGIuGIuLeuGIYLeVAlGIYSeRGIInSeRSeRPhSeRGISeRMeTg 660
Db 2077 ACGGTCACCGGAAGAGCTGGGAAGTGGGAGCTAGCTTACCTTTTCGGGACGATG 2136
Qy 660 lULeIleGIuValSeR 665
Db 2137 AATCATTTAGGCTCC 2153

RESULT 10
AX441229 3332 bp DNA linear PAT 28-JUN-2002
LOCUS AX441229 Sequence 20 from Patent WO0226997.
DEFINITION AX441229
ACCESSION AX441229
VERSION AX441229.1 GI:21665771
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Lucbe,R.M. and Wei,B.
TITLE Dep-16 dual-specificity phosphatase
JOURNAL Patent: WO 0226997-A 20 04-APR-2002;
Ceptyr, Inc. (US)
FEATURES
source location/Qualifiers
1..3332
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Alignment Scores:
Pred. No.: 0 Length: 3332
Score: 472.00 Matches: 472
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 70.98% Indels: 0
DB: 6 Gaps: 0
US-10-029-345a-109 (1-665) x AX441229 (1-3332)

Qy 194 ThrCybProLybProApHeRILeProGIuSeRHiSPhELeUaRgVALProVALAsnAp 213
Db 977 ACTGTCCAAGAGCTGACTTATCCCGAGTCTCATTTCTCGCGTGGCTGGAAATGAC 1036
Qy 214 SerPhCySGInLUaLaseRILeUProThrPLeuApLybSeRValAspHeRILeGIuLybAla 233
Db 1037 ACCTTTTGAGAAAAATTTGGCGTGGTGGCAATAGTGAATTTCAATTGGAAGA 1096
Qy 234 LybAlaSeRAsnGIYCybValLeuVALHISCybLeuVALGIYLeSeRAspSeRAlaThr 253
Db 1097 AAGGCTCCAAATGATGTCTTCTAGTGCATCTGTTAGCTGGATCTCCGCTCCGAC 1156
Qy 254 lIeAlaIleAlaTYrILeUeRlybARgWeLAspMeLSeRLeuApGIuAryrARgPh 273
Db 1157 ATCGCTATCGCTACATCATGAAGAGATGACATGTCTTTAGATGAACCTTACAGATT 1216
Qy 274 ValLybGIuLybARgProThrILeSeRProAnPhEaRPhELeUGIYGIuLeuAaP 293
Db 1217 GTGAAGAAAAAAGACTCTATCTATCTCAAACTTCATTTCTGGGCCAACTCCAGAC 1276
Qy 294 TYrGIuLybValILeLybAsnGINThrGIYAlaSeRGIYProLybSeRlybLeuYbLeu 313
Db 1277 TATGAAGAAATTAAGAACCAAGATGAGATGAGGAGCAAGAGCAAACTCAAGCTG 1336
Qy 314 LeUHiSLeUGIuLybProAnGIuProVALProAlaValSeRGIuGIYGIuLybSeR 333
Db 1337 CTGACACTGAGAGAGCAATGAACTGTCCGTGCTTCAGAGGGTGGACGAAGAAC 1396
Qy 334 GIuThrProLeuSeRProProCybAlaAspSeRAlaThrSeRGIuAlaIleGIYInARg 353
Db 1397 GAGACCCCTCAAGTCCACCTGTGCCGACTGTGCTCACTCAGAGGACAGACAGCAAAAG 1456
Qy 354 ProVALHISProAlaSeRValProSeRValProSeRValGIYProSeRLeuGIuAaP 373
Db 1457 CCGGTCATCCGCGACAGCTGCCAGCGTCCAGCGTCCAGCGTCCGCTTTAGAGAC 1516
Qy 374 SerProLeuVALGIuAlaLeuSeRGIYLeUHiSLeUSeRAlaAspARgLeuGIuAaPSeR 393
Db 1517 AGCCGCTGTGTAAGGCGCTCAAGTGGCTGCACTGTCCGAGACAGGCTGGAAAGACAG 1576
Qy 394 AsnLybLeuLybARgSeRPhSeRLeuAaPILeLybSeRValSeRTYrSeRAlaSeRMeT 413
Db 1577 AATAAGCTCAACAGTTCTTCTCTCGATTAACAATGATTTCATATTCAGCCACACAG 1636
Qy 414 AlaAlaSeRLeUHiSGIYPhSeSeRSeRGIuAaPAlaLeuGIuLYrTYrLybProSeR 433
Db 1637 GAGCATCTTACATAGCTTCTCTCATAGAAAGATCTTTGGAATACTACAAACCTTCC 1696
Qy 434 ThrThrLeuAspGIYThrAsnLybLeuCybGInPhSeRProVALGInLUeUSerGIu 453
Db 1697 ACTACTCGGATGGAGCAAGAGATGCCAGTTCTCCCGTTCAGAAATCATTCGGAG 1756
Qy 454 GIuThrProGIuThrSeRProApLybGIuGIuAlaSeRILeProLybLybLeuGINThr 473
Db 1757 CAGACTCCCGAAACCGCTCTGATAGAGAGAGAGAGAGATCCCAAGAGCTGCGAGACC 1816
Qy 474 AlaARgProSeRAspSeRGIInSeRlybARgLeUHiSSeRValARgThrSeRSeRGIY 493
Db 1817 GCGAGGCTTCAGACAGCAAGAGAGATTCGCTGCTGAAGACAGCAAGCTGGC 1876
Qy 494 ThrAlaGIuARgSeRLeuSeRProLeuHiSARgSeRGIYSeRValGIuApHeNtYr 513
Db 1877 ACCGCCCAAGAGTCCCTTTATCTCCAGTGCATCAAGAGTGGAGGCTGAGAGCAATTAC 1936

QY 514 HisThrSerPheLeuPheGlyLeuSerThrSerGlnGlnHisLeuThrIysSerAlaGly 533
Db 1937 CACACCAAGCTTCCCTTTTGGCCCTTTCCACAGCAGACAGACCTCCAGAAAGTCTGGTGG 1996
QY 534 LeuGlyLeuIysGlyIleThrIleSerAspIleLeuAlaProGlnThrSerThrProSerLeu 553
Db 1997 CTGGGCGCTTAAGGCTGGCACTCGCATATCTTGGCCCCCAGACCTTACCCCTTCCCTGG 2056
QY 554 ThrSerSerTrpTyrrPheAlaThrGluSerSerHisPheTyrrSerAlaSerAlaIleTyrr 573
Db 2057 ACCAGAGAGCTGTATTTTGGCAGAGAGTCTCCACATTTACTCTGGCTCAGGCATCTAG 2116
QY 574 GlnGlySerAlaSerIleTyrrSerAlaTyrrSerCysSerGlnLeuProThrCysGlyAspGln 593
Db 2117 GAGGAGGAGTCCAGTTCCTGCTCAGCTCAGCTGAGCTGCCCTTGGCGAGACCA 2176
QY 594 ValTyrrSerValArgArgGlnIysProSerAspArgAlaAspSerArgArgSerTrp 613
Db 2177 GTCTATTCGTGTGGCAGGGCGGAGCAAGCCAGTGAAGAGCTGACTCGCGCGAGCTGG 2236
QY 614 HisGlnGluSerProPheGlnIysGlnPheIysArgArgSerCysGlnMetGlnPheGly 633
Db 2237 CATGAAGAGAGCCCTTTGAAAGCAGTTTAAACGACAGAGCTGCCAAATGGATTGGGA 2296
QY 634 GluSerIleMetSerGlnAsnArgSerArgGlnGluLeuGlyIlyValGlySerGlnSer 653
Db 2297 GAGAGCATCATGTCAAGAACAGCTCACGGAGAGAGCTGGGAGAAAGTGGGAGCTGCT 2356
QY 654 SerPheSerGlySerMetGlnIleIleGlnValSer 665
Db 2357 AGCTTTTCGGGAGCATGGAAATCATTTAGGTCTCC 2392

RESULT 11
AX441210
LOCUS AX441210 3496 bp DNA linear PAT 28-JUN-2002
DEFINITION Sequence 1 from Patent WO0226997.
ACCESSION AX441210
VERSION AX441210.1 GI:21665766
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 Luche, R.M. and Wei, B.
TITLE Dsp-16 dual-specificity phosphatase
JOURNAL Patent: WO 0226997-A 1 04-APR-2002;
Cepcyr, Inc. (US)

FEATURES
source 1..3496
location/Qualifiers
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 3496
Score: 472.00 Matches: 663
Percent Similarity: 99.40% Conservative: 0
Best Local Similarity: 99.40% Mismatches: 2
Query Match: 70.98% Indels: 4
DB: 6 Gaps: 0

US-10-029-345A-109 (1-665) x AX441210 (1-3496)

QY 1 MetAlaHisGluMetIleGlyThrGlnIleValThrGluArgLeuValAlaLeuLeuGln 20
Db 562 ATGGCCCATGAGATGATTGGAACTCAATTTGTTACTGAGAGGTGTGGCTCTGTGTGAA 621
QY 21 SerGlyThrGlnIlyValIleLeuLeuIleAspSerArgProPheValGlnTyrrAsnThrSer 40
Db 622 AGTGAACGGAAAAAGTGTCTTAATTGATGACCGGCATTTTGGAAATACATATCATCC 681

QY 41 HisIleLeuGlnAlaIleAsnIleAsnCysSerIlyLeuMetIysArgArgLeuGln 60
Db 682 CACATTTTGAAGCCATTAAATATCAATGCTTCAAGCTTATGAAAGGATTGCAACAG 741
QY 61 AspIysValIleIleThrGlnLeuIleGlnHisSerAlaIlyHisIlyValAspIleAsp 80
Db 742 GACAAAGTTTAAATACAGAGCTCATTCAGCATTCAGGAAACATTAAGGTTGACATTTAT 801
QY 81 CysSerGlnIysValIleValTyrrAspGlnSerSerGlnAspValAlaSerLeuSerSer 100
Db 802 TGCAGTCAAGAGGTTGTATGTTTACATCAAGAGCTCCAAAGATGTTGCCCTCTCTCTTCA 861
QY 101 AspCysPheLeuThrValIleLeuGlnIlyIysLeuGlnIysSerPheAsnSerValHisLeu 120
Db 862 GACTGTTTTCACAGTACTTCTGGGTAACTGGAGAAAGACTTCAACTCTGTTCCACTGG 921
QY 121 LeuAlaGlyIlyPheAlaGluPheSerArgCysPheProGlyLeuCysGlnGlyIlySer 140
Db 922 CTTCAGAGTGGGTTTGTCTGAGTCTCTGTGTGTTTCCCTGGCCTTGTGAAGAAATTC 981
QY 141 ThrLeuValProThrCysIleSerGlnProCysLeuProValAlaAsnIleGlyProThr 160
Db 982 ACTGATGCTCCACTGCTGCTTCTCAGCCTTGCTTACTGTTGGCAACATTGGGCAAC 1041
QY 161 ArgIleLeuProAsnLeuTyrrLeuGlyCysGlnArgAspValLeuAsnIysGlnLeuIle 180
Db 1042 CGAATTTCTCCCAATCTTATCTTGGCTGACGAGAGATGCTTCACAAAGAGCTGAT - 1100
QY 181 -GlnGlnAsnGlyIleGlyTyrrValLeuAsnAlaSerTyrr-ThrCysProIysProAsp 200
Db 1101 GCAGCAGATGGATGTTGTTATGTTTAAATGCCAGCA - TACCTGTCCAAAGCTTGACT 1159
QY 200 IleIleProGluSerHisPheLeuArgValProValAsnAspSerPheCysGlnIlyIle 220
Db 1160 TTATCCCGAGTCTCATTTCTGCGTGGCTGGAATGACAGCTTTTGTGAGAAATTT 1219
QY 220 EupProTrpLeuAspIysSerValAspPheIleGlnIlyValIleAsnIlyCysVal 240
Db 1220 TGCCTGTGTGGAAATCACTAATATTTCAATTGAAAGCAAAAGCTCCCATATGAGATGTG 1279
QY 240 AlIleuValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrrIle 260
Db 1280 TTCTAGTCACTGTTTACGTGGATCTCCGCTCCGCAACATGCTATGCTTACATCA 1339
QY 260 eIlyArgMetAspMetSerLeuAspGlnAlaTyrrArgPheValIysGlnIlyAspProT 280
Db 1340 TGAAGAGATGACATGCTTTAGATGAAGCTTACAGATTTGTGAAGAAAGAAAGACTTA 1399
QY 280 hrIleSerProAsnPheAsnPheLeuGlyIleLeuLeuAspTyrrGlnIlyIleIysAla 300
Db 1400 CTATATCTCCAAACTTCAATTTCTTGGGCCCACTCTGAGCTATGAAGAAAGATTAGA 1459
QY 300 snGlnThrGlyAlaSerGlyProIysSerIlyLeuIysLeuHisIleGlnIlyAspPro 320
Db 1460 ACCAGACTGAGATTCAGGGCCAAAGCAAACTCAAGCTGCTGCACTGGAAGAGCCAA 1519
QY 320 snGlnProValProAlaValSerGlnIlyGlnIlySerGlnIlyThrProLeuSerProp 340
Db 1520 ATGAACCTGTCCTCGTGTCTCAGAGGGTGAGACGAAAGACGAGACCCCTCAGTCCAC 1579
QY 340 rocYsaIlaAspSerAlaThrSerGlnAlaIleGlyIleGlnArgProValHisProAlaSerV 360
Db 1580 CTGTGCTGACTGCTACTCAGAGGACAGAGCAAGAAAGCCCGTGCATCCCGCCAGCG 1639
QY 360 alProSerValProSerValGlnProSerLeuLeuGlnIlyAspSerProLeuValGlnAla 380
Db 1640 TGCCACAGCTGCCCAAGCTGACGCTGCTGTTAAGAGACAGCCGCTGTACAGGCGC 1699
QY 380 eusSerGlyLeuHisIleuSerAlaAspArgLeuGlnAspSerAsnIlyLeuIlyAspSerP 400
Db 1700 TCAGTGCGGTGACACTGTCCGACAGACAGGCTGGAAGACACATTAAGCTCAACGTTCC 1759

QY 400 heserleuaplleuyservalsevserlasermetalaalaserleuhsiglyp 420
 DB 1760 TCTCTGTGATATCAATCAAGTTTATATTCACCGCAGCATGGAGCATCTTACATGAGCT 1819
 QY 420 hesersersergluapalaleuGluTyrTyrlyserProserThrlleuapglYThra 440
 DB 1820 TCTCCCTCATCAGAAAGATGCTTGGATAGTACAAACCTTCCACTGATCTGATGGAGACA 1879
 QY 440 gplyleuyserglnpbeserprovalGlnGluuenserGluGlnThrProGlnThrserp 460
 DB 1880 ACAAGCTATGCGAGTTCTCCCTGTCAGAACTATCGAGCAGCATCCCGAAACCAAGTC 1939
 QY 460 roasplysglugluajalaserleProlylsyleruGlnThralaargProseraspserg 480
 DB 1940 CTGATAGAGAGAGAGCAGCATCCCAAGAGCTGAGAGACCGCAGGCTTCAGAGAGCC 1999
 QY 480 inserlyargleuhsiservalarThrserserserglyThralaglnarsergleu 500
 DB 2000 AGAGCAAGCGATTCATTCGGTCAAGAACGAGCAGCATGGCAGCCCGCAGAGGTCCTTT 2059
 QY 500 euserProleuhsisarserglyserValGluasparlyThrlthrserpheleuphag 520
 DB 2060 TATCTCCTCAGTCAATCGAAGGGAGCGTGAAGCAATTACACACAGCTTCTCTTTG 2119
 QY 520 lyLeuserThrserserglnGlnhsleuThrlserseralaglyleuGlyleuGlyLTPH 540
 DB 2120 GCTTTTCCACACCGCAGCAGCAGCTTCAAGAGCTCTGCGCTTGAAGGCTTGGC 2179
 QY 540 lsseraplleleualaproGlnThrsThrsProserleuThrsersertPyrPhea 560
 DB 2180 ACTCGATATCTTGGCCCCAGACCTCAACCTTCCCTGACACAGCTGATTTTG 2239
 QY 560 lathrluserserhsisepheThrserseralaleThrglylyserlaserlyrs 580
 DB 2240 CCACAGAGCTTCAACCTTCTACTCTGCTCAGCCACTTCAAGAGCAGCTGAGCTTACT 2299
 QY 580 eeralatyserserserglnleuProThrsCysaglyleuaplnValTyservalararxa 600
 DB 2300 CTGCTTACAGCTGACAGCAGCTCCCACTTGGAGAACCAAGCTTCTTCTGCGCAGGC 2359
 QY 600 rsglnlyarProserasparalalaspserarararserThrlhaglugluserProphag 620
 DB 2360 GGCAGAGAGCAAGTGAAGAGCTGACTCGCGCGGAGCTGAGTGAAGAGAGCCCTTTG 2419
 QY 620 lulyeglnpheylararararserCysglmeteluehlygluserllemetserglun 640
 DB 2420 AAAAGAGCTTAAACCGACAGAGCTGCCAATGGAATTTGAGAGACATCATGTGAGAGA 2479
 QY 640 sarysersar 660
 DB 2480 ACGAGTCAAGAGAGCTGGGAGAGAGTGGGAGTCAAGTCTTTCGGGAGCAGCATG 2539
 QY 660 lulleleugluvalser 665
 DB 2540 AATCATTTGAGTCTCC 2556
 DB
 RESULT 12
 AB052156 3521 bp mRNA linear PRI 18-OCT-2001
 LOCUS Homo sapiens MKP-7 mRNA for MAPK phosphatase-7, complete cds.
 DEFINITION
 AB052156
 ACCESSION
 VERSION
 AB052156.1 GI:13548676
 KEYWORDS
 SOURCE
 Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE
 1 Masuda, K., Shima, H., Watanabe, M., and Kikuchi, K.
 TITLE
 MKP-7, a novel mitogen-activated protein kinase phosphatase,
 functions as a shuttle protein
 JOURNAL
 J. Biol. Chem. 276 (42), 39002-39011 (2001)
 MEDLINE
 21486429

PUBMED 11489891
 2 (bases 1 to 3521)
 REFERENCE
 AUTHORS Masuda, K., Shima, H., and Kikuchi, K.
 TITLE Direct Submission
 JOURNAL Submitted (05-DEC-2000) Kouhei Masuda, Institute for Genetic
 Medicine, Division of Biochemical Oncology and Immunology, Kita-ku
 Kitas-5-jou nishi17-cyome, Sapporo, Hokkaido 060-0815, Japan
 (E-mail: kou@imm.hokudai.ac.jp, Tel: 81-11-706-5536,
 Fax: 81-11-707-6839)
 FEATURES
 source
 1. 3521
 location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /chromosome="12"
 /cell_line="Jurkat"
 1. 3521
 /gene="MKP-7"
 564. 2561
 /gene="MKP-7"
 /codon_start=1
 /product="MAPK phosphatase-7"
 /protein_id="BAB40814.1"
 /db_xref="GI:13548677"
 /translation="MAHEMIGTQVTERLVALLSEGTKEVLLIDSRPFVEVNTSHILE
 ALINCSKLMKRRLOODKVLITELIOHSAKHKVVDIDCSOKVVVYVSDVASLSDC
 FLTVILGLKESFNSVHLIAGFAFSPKCFELCEKSTIVPTCSIOPELPIANTGPT
 RILPNLYGCGQDVNLKELMOONGIGVYLNASNTCPKPDFIPESHFLRPVVDSCFEK
 ILPLWDSKVDPIEKAKASNGCVLHCLAGISASATIALAYIKRMDSIDEXYRVEKE
 KRPTSPNPNFLGQLDYEKIKNOTGASGPKSKLHLKEPNEPVPVSVSGQKSE
 TPLSPCADSATSEAAAGORPVHPASVPSPVLPLEDSPLVQALSGHLSADRLRE
 SNKLRSPLSDIKYSVSASMASTHSGSSSPDLEFYKPSFTTLGCTNLCOFSPVQE
 LSEQPTSPDKEASTIRKQUTARPSPSQSRHLISVTRSSSGTRQSRILSLHNSGS
 VEDNHTFLFGLSTSQHLISAGLAKMHSIDLAEPTSPSLTSVYFATESHF
 YSAAIYGSFASVSQSOPLTCGDQVSVARRKQPSDRADSRSSWHESPFEKQFK
 RSSCOMEGESIMSENREBELKVGSSSFEFGSMELIEVS"
 ORIGIN
 Alignment Scores:
 Pred. No.: 0 Length: 3521
 Score: 472.00 Matches: 663
 Percent Similarity: 99.40% Conservative: 0
 Best Local Similarity: 99.40% Mismatches: 2
 Query Match: 70.98% Indels: 4
 DB: 9 Gaps: 0
 US-10-029-345A-109 (1-665) x AB052156 (1-3521)
 QY 1 MetalaHsagluMetlleglyThrglnleValThrgluarGluValalaleuGlu 20
 DB 564 ATGGCCCATGAGATGATGGAACTCAAAATTTGACTGAGAGGTGGCTGCTGGAGAA 623
 QY 21 SerGlyThrgluValleuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 40
 DB 624 AGTGAAGCGAAGAAAGTGTCTATTTGATATACCGGCACTTGTGGAATACATACATCC 683
 QY 41 HsalleleuGluAlalleuhsanlleuhsanlleuhsanlleuhsanlleuhsanlleu 60
 DB 684 CACATTTTGGAGCATTATATCAACGCTCCAGCTTATAGAGGAGGATGGCAACAG 743
 QY 61 AsplyvalleuLeuThrgluLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 80
 DB 744 GACAAAGTTTAAATTCAGAGCTCATCCGATTCGCCAAACATTAAGATTGACATTGAT 803
 QY 81 CyserserglnlyValValValValValValValValValValValValValValVal 100
 DB 804 TCGAGTCAGAAAGTTGTGATTACATCAAAAGCTCCAAAGAGTGGCTCTCTCTTCA 863
 QY 101 AapCyspheleuThValleuLeuGlylylsleuGlylylsleuPheamserValhsleu 120
 DB 864 GACTGTTTTCTCAGCTGATCTTCTGGGTAAACGAGAAAGCTTCAACTCTGTCACCTG 923
 QY 121 LeuAlaglyGlyPhealagluPheaserarGyspheProGlyLeuCyagluglylyrs 140

```

Db      924  CTTGCAAGGTGGTTGTGTAAGTTCTCTCGTTGTTCCCTGCTCTGTGAAGAAATTC 983
Qy      141  ThrLeuValProThrCysIleSerGlnProCysLeuProValAlaAsnIleGlyProThr 160
Db      984  ACTCTAGTCCCTTACCTGTCATCTTCTGACCTTGTCTTACCTGTCACCAATGGGGCAACC 1043
Qy      161  ArgIleLeuProAsnLeuTyrLeuGlyCysGlnArgAspValLeuAsnIleGlyLeuIle 180
Db      1044  CGAATCTTCCCAATCTTATCTTGGCTGTCAGAGATGTCTTCAACAAAGAGCTGAT- 1102
Qy      181  -GlnGlnAsnGlyIleGlyTyrValLeuAsnIleAspTyr-ThrCysProIleProAsp 200
Db      1103  GCAGCAAGATGGATGGATTGTTATGTTAAATGCCAGAA-TACTCTTCCAAAGCTGACT 1161
Qy      200  helLeuProGlnSerHisPheLeuArgValProValAsnAspSerPheCysGlnIleLeu 220
Db      1162  TTATCCCCGAGTCTCATTTCTCTGCTGCTGTGATGACAGCTTTTGTGAAATAATT 1221
Qy      220  euProTribLeuAspLysSerValAspPheIleGlnIleValAlaSerAsnGlyCysVal 240
Db      1222  TGGCGTGGTTGGCAAAATCATGATTTTCAITGAGAAAGCAAAAGCTCCAAATGATGTCG 1281
Qy      240  alLeuValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIleAl 260
Db      1282  TTCTAGTGCACTGTTAGCTGGGATCTCCGCTCCGCAACATCGCTATCGCTCATCATCA 1341
Qy      260  euLysArgMetAspMetSerLeuAspGlnAlaTyrThrPheValIleGlyValArgProT 280
Db      1342  TGAAGAGATGACATGCTTTAGATGAAAGCTTACAGATTTGTGAAAGAAAAAGCTTA 1401
Qy      280  hrIleSerProAsnPheAsnPheLeuGlyGlnLeuAspTyrGlnIleValIleVal 300
Db      1402  CTATATCTCCAAATCTCAATTTTCTGGGCCAACTCTGAGCATHTGAGAAAGATTAAGA 1461
Qy      300  snGlnThrGlyAlaSerGlyProLysSerLysLeuLysLeuHisLeuGlnIleValPro 320
Db      1462  ACCAGACTGAGATCATGGGCCCAAGAGCAAACTCAAGCTGTCGACCTGGAGAAAGCCA 1521
Qy      320  snGlnIleProAlaValSerGlnIleGlyGlnIleSerGlnIleThrProLeuSerProP 340
Db      1522  ATGAACTGTCTCTGTCTGTCAGAGGGTGAAGAGAAAGCCAGAGCCCTTCACTTCCAC 1581
Qy      340  roCysAlaAspSerAlaThrSerGlnAlaAlaGlyGlnArgProValHisProIleSerV 360
Db      1582  CCTGTGCCGACTCTGTACTCATAGAGGACAGAGCAAAAGCCCTGATCCCGCCAGCG 1641
Qy      360  alProSerValProSerValGlnProSerLeuLeuGlnAspSerProLeuValGlnAla 380
Db      1642  TCCCCAGCGTCCCGCAGCGTCCGCTGTTAGAGAGACAGCCGCTGTGAACAGGCGC 1701
Qy      380  euSerGlyLeuHisLeuSerAlaAspArgLeuGlnAspSerAsnIleLeuValAspSerP 400
Db      1702  TTAGTGGGCTGCACTGTCCGAGACAGGCTGGAAGACGCAATAGCTCAAGCCCTTCT 1761
Qy      400  heSerLeuAspIleLysSerValSerTyrSerAlaSerMetAlaIleSerLeuHisGlyP 420
Db      1762  TCTCTCTGGAATCAATCAGTTTCAATTCAGCCAGACATGSCACATCTTACATGGCT 1821
Qy      420  heSerSerSerGlnAspAlaLeuGlnIleTyrLysProSerThrThrLeuAspGlyIle 440
Db      1822  TCTCTCTCATCAAAATGCTTGTGAATCTCAAAATCTTCCACTACTCTGAGTGGAGCA 1881
Qy      440  snLysLeuCysGlnPheSerProValGlnIleLeuSerGlnIleThrProGlnIleThr 460
Db      1882  ACAAATGTCAGAGTTCCTCCCTGTTCAAGAACTATCGAAGAGCTCCCAAAACCAATC 1941
Qy      460  roAspLysGlnGlnAlaSerIleProLysLeuGlnIleThrAlaArgProSerAspSerG 480
Db      1942  CTGATTAAGAGAGAACCAAGCATCCCAAGAGCTGCAACGCGCAGGCTTCAAGACGCC 2001
Qy      480  IAspLysArgLeuHisSerValArgThrSerSerSerGlyThrAlaGlnArgSerLeu 500

```

```

Db      2002  AGAGCAAGCATTCGATTCGTCAGAACACAGACGATGGCACCGCCCAAGATTCCTTT 2061
Qy      500  euSerProLeuHisArgSerGlySerValGlnAspAsnTyrHisThrSerPheLeuPheG 520
Db      2062  TATCTCCACTGCATGAGAGTGGAGAGCTGGAGAGCAATTACACACAGACTTCTTTGCG 2121
Qy      520  ILeuSerThrSerGlnGlnIleHisLeuThrLysSerAlaGlyLeuGlyLeuIleGly 540
Db      2122  GCTTTTCCACAGCCAGACGACCTCAAGAGTGTGCTGGCTGGCTTGAAGGCTGGC 2181
Qy      540  IAspAspIleLeuAlaProGlnIleThrSerThrProSerLeuThrSerSerTyrPheA 560
Db      2182  ACTGGATATCTGGCCCCCAGACCTTACCTTCCCTACACAGACGCTGATTTTG 2241
Qy      560  IArgHisSerSerHisPheTyrSerAlaSerAlaIleTyrGlyCysSerAlaSerTyrS 580
Db      2242  CCACAGAGTCCCTCACACTTACTCTGCTGACGACATCTAGGAGGAGCTCCAGTTACT 2301
Qy      580  erAlaTyrSerCysSerGlnIleuProThrCysGlyAspGlnValTyrSerValArgArg 600
Db      2302  CTGCTTACAGCTGACGCGCTGCCACTTGGAGAGACCAAGTCTATCTGTGGCGGC 2361
Qy      600  rGlnIleProSerAspArgAlaAspSerArgArgSerTyrPheIleGlyGlyIleProPheG 620
Db      2362  GCGAGAAACCAAGTACAGAGACTGACTGCGCGGAGCTGGCATGAAAGAGCCCTTTG 2421
Qy      620  ILeuGlnPheLysArgArgSerCysGlnIleGlnIleGlyIleGlyIleMetSerGln 640
Db      2422  AAAAGCATTTTAAACGCAAGAGCTGCCAAATGAAATTTGGAGAGACATCATGCAAGA 2481
Qy      640  snArgSerArgGlnGlyLeuGlyLysValGlySerGlnIleSerPheSerGlySerMetG 660
Db      2482  ACAAGTCAAGGAGAGAGCTGGGAGAAAGTGGACAGCTAGCTTATGCTTGGAGCAATGG 2541
Qy      660  IuileIleGlnValSer 665
Db      2542  AAATCATTAGAGTCTCC 2558

RESULT 13
AX260340
LOCUS      3544 bp      DNA      linear      PAT 26-OCT-2001
DEFINITION Sequence 1 from Patent WO0173059.
ACCESSION  AX260340
VERSION     AX260340.1 GI:16509303
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE   1
            Meyers, R.A.
            38692 and 21117: dual specificity phosphatase molecules and uses
            thereof
            Patent: WO 0173059-A 1 04-OCT-2001;
            Millennium Pharmaceuticals, Inc. (US)
FEATURES
            source
            location/Qualifiers
                1..3544
                /organism="Homo sapiens"
                /mol_type="unassigned DNA"
                /db_xref="taxon:9606"
                589..2586
                /note="unamed protein product"
                /codon_start=1
                /protein_id="CAD10218.1"
                /db_xref="GI:16509304"
                /db_xref="REMTREMBL:CAD10218"
                /translation="MAHEMIGQIVTERIVALLSESTEKVLIDSRPFVENVTSIIE
                AININSGKMKRRLOODKYLITELIOHSAKHVDIDCSQKVVYVQDSODVASISDC
                RITPLVGLKLEKSNHYHLAAGFAESRFRPGICEBKSTLVPICISQPLPIANISPT
                ILFWIDKSVDFLEKAVASNGCVLVHGLAISRTATLAIYIKRMDSIDEAYRVFKE
                KRPTISPNFNFLQQLIDYEKTIKNQGTGAGSPKSKLILHLEKPNFVPAVSGGQKSE
                TPLSPCADSATSEAGRPVHPASVPSVQPSILBDSPIVQALSGHLISADRLIED

```

ORIGIN

SNLKRSFSLDIKSVSYASAMASLHGFSSSEDALEYKSPITLTDINKLCQFSPVCE
LSEQNPETSPDKEBASIPKILQIARPSDSQKRLHSVRTRSSGTSORSLISPLHSGS
VEDNYTSLFRLSLSTSOOLHTYSAGLGLKMSHDLARPOTSPSLTSSRYFTSHSF
YASASITGGSASYSATSCQLPICQDQVTSRRKQPSRRAQSRKSMHSESPFERQPK
RSCQMEFSESIMSENRSERELKGVSSQSSFSGSMELIEVS"

Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
0	472.00	99.40%	99.40%	70.98%	3544	663	0	2	4	0

US-10-029-345A-109 (1-665) x AX260340 (1-3544)

QY 1 MetAlaHiegluMetileglThrgluHleValThrgluArgLeuValAlaLeuLeuGlu 20
Db ATGGCCCATGAGATGATGGAATTCMAATTTGTAAGAGATGGTGGCTGTGCGAA 648
QY 21 SerGlyThrgluLeuValLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 40
Db AGTGAACCGGAAAGAGTGGTGAATGATGACCGGCAATTTGTGGAATACATACATCC 708
QY 41 HselleLeuGluAlaHleLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 60
Db CACATTTTGAAAGCCATTAATATCACTGCTCCACCTTAATGAGGAAAGTTGCAACG 768
QY 61 AspLeuValLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 80
Db GACAAAGTGTAAATTCAGAGCTCATCCAGCATTCAGCAACATTAAGTTGACATGAT 828
QY 81 CysSerGluLeuValValValValValValValValValValValValValValVal 100
Db TGCAGTCAGAGAGTGTAGTTACGATCAAGCTCCAGAGATGTTCTCTCTCTCA 888
QY 101 AspCysPheLeuThrValLeuLeuLeuGlyValLeuGluLeuSerPheLeuValHleu 120
Db 889 GACTGTTTCTCAGCTGACTCTGGGGTAACTGGAGAAAGCTTCAACTCTGTTCACTG 948
QY 121 LeuAlaGlyGlyPheAlaGluPheSerArgCysPheProGlyLeuGlyGlyLeuSer 140
Db 949 CTTCGAGGTGGTTCCTGAGTCTCTCTGTTTCTGCTGCTCTGTAAGAAATATC 1008
QY 141 ThrLeuValProThrCysIleSerGlnProCysLeuProValAlaLeuLeuLeuLeu 160
Db 1009 ACTCTAGTCCCTACCTGCAATTTCTCAGCTTGTCTTACTGTTGCCAAATTGGGCCAACCC 1068
QY 161 ArgIleLeuProAlaLeuLeuLeuLeuGlyCysGlnArgAspValLeuLeuLeuLeuLeu 180
Db 1069 CGAATTTCTTCCCAATCTTAATCTTGGCTCCAGCGAGATGCTCTCAACAGAGCTGAT- 1127
QY 181 -GlnGlnAanglyIleGlyTyValLeuLeuAlaSerTyr-ThrCysProLeuProAsp 200
Db 1128 GCAAGCAATGGATGGTGTATGTATGTAATGCCAGCA-TACCTGTCCAAAGCCGAGCT 1186
QY 200 HselleProGluSerHisPheLeuArgValProValAlaLeuLeuLeuLeuLeuLeu 220
Db 1187 TTAATCCCGAGTCTCAATTCCTGCTGCTGCTGTAATGACAGCTTTGTGAGAAATATT 1246
QY 220 eulProThrLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 240
Db 1247 TGCCTGGTGGTGAACATAGATTTCAATGAGAAAGCAAGCTTCAATGAGATGAG 1306
QY 240 alLeuValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyIleAl 260
Db 1307 TTCTAGTCACTTTTAACTGAGATCTCCGCTCCCGCCACATCGCTTACGCTTACATCA 1366
QY 260 eulValArgMetSerLeuLeuAlaValValValValValValValValValValVal 280
Db 1367 TGAAGAGATGACATGTCTTTAGATGAGCTTACAGATTTGTGAGAAAGAAAGAGCTTA 1426

QY 280 HselleSerProAlaLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 300
Db 1427 CTAATCTTCCAACTTCAATTTTCTGGGCAACTCTGAGCAATAGAGAAAGATTAA 1486
QY 300 eulGlnThrGlyAlaSerGlyProLeuSerLeuLeuLeuLeuLeuLeuLeuLeuLeu 320
Db 1487 ACCAGACTGAGCATCATGGGCAAGAGCAAACTCAAGCTGCTGCACTGAGAGAGCCAA 1546
QY 320 eulGluProValProAlaValSerGluGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 340
Db 1547 ATGAACCTGTCT 1606
QY 340 roCysAlaAspSerAlaThrSerGluAlaAlaGlyGlnArgProValHisProAlaSerV 360
Db 1607 CTTGTGCGACT 1666
QY 360 alProSerValProSerValGlnProSerLeuLeuLeuLeuLeuLeuLeuLeuLeu 380
Db 1667 TCCAGAGGTGCGCCAGCGGCGAGCCCTGCTGTAGAGACAGCCGCTGTACAGAGCC 1726
QY 380 eulSerGlyLeuHleLeuSerAlaAspArgLeuGluAspSerLeuLeuLeuLeuLeu 400
Db 1727 TCAGTGGGCTGCACCTGTCCGAGACAGCTGGAAGACAGCAATAGCTCAAGCCGTTCT 1786
QY 400 HselleLeuAlaIleValSerValSerTyrSerAlaSerMetAlaAlaSerLeuHleGly 420
Db 1787 TCTCTCTGATATCAATATGATTTATTTACGCCAGATGAGCAGATCTTACATGCT 1846
QY 420 HselleSerSerGluAlaLeuGluTyTyTyTyTyTyTyTyTyTyTyTyTyTyTyTy 440
Db 1847 TCTCTCTCAACAGAGTCTTTGGAATACTAACAACTTCCACTGCTGATGGAGCA 1906
QY 440 eulLeuLeuCysGlnPheSerProValGlnGluLeuSerGluGlnThrProGluThrSer 460
Db 1907 ACNACCTATGCGAGTCT 1966
QY 460 roAspLeuGluGluAlaSerIleProValValLeuGlnThrAlaArgProAspArgSer 480
Db 1967 CTGATTAAGAGAGAGCCAGCATCCCAAGACCTGACAGCCCGCTTACAGAGCC 2026
QY 480 InsleuValArgLeuHleSerValArgThrSerSerSerGlyThrAlaGlnArgSerLeu 500
Db 2027 AGAGCAGAGATGATTCGATTCGACAGACAGAGAGAGAGAGAGAGAGAGAGAGAG 2086
QY 500 eulSerProLeuHleArgSerGlySerValGluAspAlaThrHisThrSerPheLeuPhe 520
Db 2087 TATCTTCCACTGCATGAGAGTGGAGAGCTGGAGAGCAATTAACAACAGCTTCTTTTG 2146
QY 520 IlyLeuSerThrSerGlnGlnHleLeuThrValSerAlaGlyLeuGlyLeuValGlyTy 540
Db 2147 GCTTTTCCAGCCAGCAGCACTTCAAGATCTGCTGCTGCTGCTGCTGCTGCTGCTG 2206
QY 540 IseSerAspIleLeuAlaProGlnThrSerThrProSerLeuThrSerSerTyrPheAl 560
Db 2207 ACTCGGATATCTTGGCCCCCAGACCTTACCTTCCCTGACAGAGCTGATTTTG 2266
QY 560 IatThrGluSerSerHisPheTySerAlaSerAlaIleTyGlyGlySerAlaSerTyrS 580
Db 2267 CCACAGAGCTCCACACATTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2326
QY 580 eulAlaTySerCysSerGlnLeuProThrCysGlyAspGlnValTySerValArgArg 600
Db 2327 CTGCTTACAGCTGAGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2386
QY 600 rglLeuLeuProSerAspArgAlaAspSerArgArgSerTyrHisGluGluLeuProPhe 620
Db 2387 GGCAGAGCAAGTACAGAGCTGACCTGCGCGAGAGCTGAGATGAAGAGAGAGCCCTTG 2446
QY 620 IlyLeuGlnPheValArgArgSerCysGlnMetGluPheGlyGluSerIleMetSerGlu 640
Db 2447 AAAACAGATTTAAAGCAGAGAGCTCCAAATGAAATTTGGAGAGAGCATCAATGATGA 2506
QY 640 snArgSerArgGluGluLeuGlyValGlySerGlnSerSerPheSerGlySerMetG 660

|||||
Db 2507 ACAGGTACCGGAAGACTGGGAAATGGGCACTGACTTTCGGGACGATGG 2566
QY 660 lullelelguValSer 665
Db 2567 AAATCATTTGAGTCTCC 2583
RESULT 14
AF506796 3566 bp mRNA linear PRI 30-OCT-2003
LOCUS AF506796
DEFINITION Homo sapiens dual specificity phosphatase 16 (DUSP16) mRNA,
complete cds.
AF506796
VERSION AF506796.1 GI:25573087
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 3566)
AUTHORS Hoornaert, I., Marynen, P., Goris, J., Sciot, R. and Baens, M.
TITLE MAPK phosphatase DUSP16/MKP-7, a candidate tumor suppressor for
chromosome region 12p12-13, reduces BCR-ABL-induced transformation
JOURNAL Oncogene 22 (49), 7728-7736 (2003)
PUBMED 14586399
2 (bases 1 to 3566)
REFERENCE Hoornaert, I., Marynen, P. and Baens, M.
AUTHORS Direct Submission
JOURNAL Submitted (26-APR-2002) Department for Human Genetics-Flanders
Interuniversity Institute for Biotechnology (VIB), Katholieke
Universiteit Leuven, Herestraat 49, Leuven B-3000, Belgium
FEATURES
source
1.3566
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="12"
/map="12p13-p12"
/cissue_type="bone marrow"
1.3566
/gene="DUSP16"
/note="synonym: MKP7"
633..2630
/gene="DUSP16"
/EC_number="3.1.3.48"
/EC_number="3.1.3.16"
/codon_start=1
/product="dual specificity phosphatase 16"
/protein_id="AAW75120.1"
/translation="MAHEMIQTQIYTERLVALLSEGTREKVLIDSRPVEVNTSHILE
AININCSKLMKRLQODKVLITELIOHSAKHEKVIDGSOXVVYDSSODVASISDPC
FTVLVIGLKEKSPNSVHLAAGFAEFSSCPGLCGKSGTIVPTCSOPCLPVANIGPT
RLPVLVIGCCQDVVNLKELMOONGI GYLVNANSTCPKDFIPESHFLVPPVNDSECEK
KRPITSPENFVGLDLYBKIKNGTGAAGPSKIKLALHEKPEPVAIVEGQKSE
TLPSPCCSDATSENAQORPVHPSAVPSPVSPLEDSPLVQALSGIHSADRLD
SKLTKRSESLDLSVSAWMAHLHGSSSEDALEYKPSFTLDGTKLCOFSPVOE
LSEOTPEPSPDKSEASIPKIKQTARPSDQSKRLHSVATSSGTAORSLGLPARGSG
VEDNYHTSPRLGELSTSOOHLTKSAGLIGKMHSDTLAQTSTPSTSSVATATSSHP
YSASAIYGGASATYSCSQLPCTCGDYVSRRKQKPSDRADSRSRMHEESPFEKQFK
RSCOMEFGESINSENRSRELEIGKVSQSFSGSMELIEVS"

Alignment Scores:
Pred. No.: 0 Length: 3566
Score: 472.00 Matches: 633
Percent Similarity: 99.40% Conservative: 0
Best Local Similarity: 99.40% Mismatches: 2
Query Match: 70.98% Indels: 4
DB: 9 Gaps: 0

US-10-029-345A-109 (1-665) x AF506796 (1-3566)

QY 1 MetAlahIsgIuMeTIlGlyThrGlnIleValIthrGluValAlaLeuLeuGlu 20
Db 633 ATGGCCCATGATGATATTGGAACCAATTTGATCGAGGTGGTGGCTGTGGA 692
QY 21 SerGIYthrgluValLeuLeuIleAspSerArgProPheValGluTyrAsnThrSer 40
Db 693 AGTGAAGCGAAGAAAGTGTCTGCTAATTATATAGCCGCAATTTGGAAATACATACATCC 752
QY 41 HisIleleGuValAlaIleAsnIleAsnCySerIysLeuMetLysArgLeuGlnGln 60
Db 753 CACATTTTGAAGACCATTAATATCACTGCTCAAGCTTATGAGGCGAAGTTGGAACAG 812
QY 61 AspLysValLeuIleThrGluLeuIleGlnHisSerAlaValGHSlyValAspIleAsp 80
Db 813 GACAAAGTTAAATTACAGAGCTCATCAGCATTCAGGAAACATTAAGTTGACATTGAT 872
QY 81 CysSerGlnLysValIleValIleTyrAspGlnSerSerGlnAspValAlaSerLeuSerSer 100
Db 873 TCGAGTCGAAAGCTTTGTTAGTACATCAAGCTCCCAAGATGTTGCTCTCTCTTCA 932
QY 101 AspCysPheLeuThrValLeuLeuGluLysLeuGluLysSerPheAsnSerValHisLeu 120
Db 933 GACTGTTTTCACGTGACTTCTGGGTAAACTGAGAGAAAGACTTCACTGTTCACCTG 992
QY 121 LeuIleGlyGlyPheAlaGluPheSerArgCysPheProGlyLeuCyGluGlyLysSer 140
Db 993 CTGGAGGTGGGTTTGTGCAATCTCTGTTGTTCCCTGCTCTGTGAAGAAATTC 1052
QY 141 ThrLeuValProThrCysIleSerGlnProCysLeuProValAlaAsnIleGlyProThr 160
Db 1053 ACTGATGCTTACTGATGATTTCCAGCTTGCTTACTGTCGCAATTGGGCCCAACC 1112
QY 161 ArgIleLeuProAsnLeuTyrLeuGlyCysGlnArgAspValLeuAsnLysGluLeuIle 180
Db 1113 CGAATCTTCCCAATCTTAACTTGTGCTGCAGAGATGTCCTCAAGAGAGCTGAT- 1171
QY 181 -GlnGlnAsnGlyIleGlyTyrValLeuAsnAlaSerTyr-ThrCysProLysProAsp 200
Db 1172 GCAGCAGATGGATGGATGTTATGTTAAATGCCAGCA-TACTGTCCAAAGCTTACT 1230
QY 200 heIleProGluSerHisPheLeuArgValProValaAsnAspSerPheCysGluLysIleIle 220
Db 1231 TTATCCCGGAGTGCATTTCTGCTGCTGCTGATGACAGCTTTTGTGAGAAATTT 1290
QY 220 euPTrOTPLeuAspLysSerValAspPheIleGluLysAlaLysAlaSerAsnGlyCys 240
Db 1291 TGCGGTGGTGGCAAAATCAGTAGATTTCATTGAGAAAGCAAAAGCTTCAATGAGATGTG 1350
QY 240 alleuValHisCysLeuValAGIYIleSerArgSerAlaThrIleAlaIleTyrIleIle 260
Db 1351 TTCTAGTGCAGCTGTTTACGCTGGATCTCCGCTCCGCCACCAATGCTATGCTTACATCA 1410
QY 260 eLysArgMetAspMetSerLeuAspGluAlaTyrArgPheValLysGluLysArgProT 280
Db 1411 TGAAGAGATGAGACATGCTTTAGATTAACACTTACAGATTGTGAAAGAAAGAAAGACCTTA 1470
QY 280 hrIleSerProAsnPheAsnPheLeuGluLysIleLeuAspTyrGluLysLysIleLys 300
Db 1471 CTATATCTCAAACTTCAATTTCTGGGCCCACTCTCGACATATGAGAAATTAAGA 1530
QY 300 snGlnThrGlyAlaSerGlyProLysSerLysLeuLysLeuLysIleGluLysPro 320
Db 1531 ACCAGACTGAGACATCAGGCGCAAGAGCAAACTCAAGCTGCTGCACCTGGAAGAGCCAA 1590
QY 320 snGluProValProAlaValSerGluGlyGlyLysSerGluThrProLeuSerProP 340
Db 1591 ATGAACCTGTCCTGCTGCTCAGAGGTGACAGAAAGCGAGAGCCCTCAGTCCAC 1650
QY 340 rocCysIleAspSerAlaThrSerGluAlaIleGlyGlnArgProValHisProAlaSer 360
Db 1651 CTTGTGCGACTGCTGACTCTCAGAGCGAGCAAGCAAAAGGCCCGTGCATCCGCGCAGCG 1710

```

QY 360 aPProSeValPProSeValGlnPProSeLeuLeuGlnAaspPProLeuValGlnAlaL 380
DB 1711 TGCCCGAGCGTGGCCAGCGTGGAGCCCTCGCTGTAAAGACAGCCCGCTGGTACAGGCC 1770
QY 380 euSeSerGlyLeuHsLeuSeSerAlaAaspArgLeuGlnAaspSerHsLeuLysArgSerP 400
DB 1771 TCAGTGGGCTGCACCTGCTCGCAGACAGAGCTGGAAACAGCAATAGCTCAAGCGTTCT 1830
QY 400 hSeSerLeuAaspLeuLeuSeSerValSerTyrSerLysSerLysMetAlaAlaSerLeuHsGlyP 420
DB 1831 TCTCTGTGAATATCAATACATGTTTCATATTCACCGAGCATGGACGATCTTCAATGAGCT 1890
QY 420 hSeSerSeSerGlnAaspAlaLeuGlnLysTyrTyrSerProSeThrThrLeuAaspGlyTha 440
DB 1891 TCTCTCATACAGAGATGCTTGGATTAATACTCAAACTTCCACTTCTGATGGAGCA 1950
QY 440 hNylLeuCySerGlnPheSerProValGlnGlnLeuSeSerGlnGlnThrProGlnThrSerP 460
DB 1951 ACAGGTATGCCAGCTTCTCCCTGTTCAAGAACTATCGAGAGCACTCCGAAACAGTC 2010
QY 460 roAapLyseGlnGlnAlaSerLysProlYylLeuGlnThrAlaArgProSeArgSerG 480
DB 2011 CTGATTAAGAGAGAGAGCAGACATCCCAAGAGCTGACAGACCGCAGGCTTCAGACAGCC 2070
QY 480 InSeLySerArgLeuHsSerValArgThrSeSerSeSerGlyThrAlaGlnArgSerLeuL 500
DB 2071 AAGNAGAGGATTCCTGCTGCTGACAGACAGAGCTGACAGCGCCAGAGGCTCTT 2130
QY 500 euSePProLeuHsArgSerGlySerValGlnAaspAntYrHsArgSerPheLeuPheG 520
DB 2131 TATCTCATGTCATCGAAGTGGAGCGTGGAGAGCAATTAACAGACCACTTCTTTCG 2190
QY 520 YLeuSerThrSerGlnGlnHsLeuThrLysSerAlaGlyLeuGlyLeuLysGlyTyrH 540
DB 2191 GCTTTCACCGCCAGCCAGACACTCTCAAGAGCTGCTGCGCTTAAAGGCTGAGC 2250
QY 540 iSeSerAapLysLeuAlaPProGlnThrSerThrProSeLeuThrSerSerTyrPheA 560
DB 2251 ACTCGATATCTTGGCCCCCGCCAGACTTCACTTCCCTGACACAGAGTGGATTTTG 2310
QY 560 iAThGlnSeSerHsArgSerGlnLeuProThrCysGlnAaspGlnValLysSerValArgA 580
DB 2311 CCACAGAGCTCTCACTTCTTACTGCTGAGCCATCTTAACGAGGCAATGCGCACTT 2370
QY 580 eAThArgSerCysSerGlnLeuProThrCysGlnAaspGlnValLysSerValArgA 600
DB 2371 CTGCCCTACAGCTGACAGCCAGCTGCCCACTTGGGAGACCAAGCTATCTTGTGCGCAGC 2430
QY 600 iGGLNlybProSeArgAaspArgAaspSerArgArgSerTrpHsGlnGlnLeuSerProPheG 620
DB 2431 GGCAGAGCCAGATGACAGAGCTGAGCTCGCGGGGAGCTGGCAGTGAAGAGAGCCCTTG 2490
QY 620 iLylLeuGlnPheLysArgSerCysGlnMetGlnPheGlyGlnSerLysMetSerGlnA 640
DB 2491 AAAAGAGCTTAAACCCAGAGCTGCCAATGGAATTTGAGAGAGCAATGTCAGAGA 2550
QY 640 sNArgSerArgGlnGlnLeuGlyLysValGlySerGlnSerPheSerGlySerMetG 660
DB 2551 ACAGGTACAGCGAGAGAGCTGGGGAAAGTGGGAGTCACTAGCTTTTGGGAGAGATGG 2610
QY 660 iLylLeuGlnValSer 665
DB 2611 AATATCATTAAGGTCTCC 2627

```

```

RESULT 15
AX374994 3766 bp DNA linear PART 01-MAR-2002
LOCUS Sequence 17 from Patent WO0210363.
DEFINITION AX374994
ACCESSION AX374994
VERSION AX374994.1 GI:19169826
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

```

```

REFERENCE
AUTHORS
Tang, Y. T., Elliott, V. S., Ramkumar, J., Yao, M. G., Burford, N.,
Wang, Y. E., Stewart, E. A., Gandhi, A. R., Patterson, C., Lee, E. A.,
Hafalia, A. J., Lu, D. A., Trounley, C. M., Griffin, J. A., Baughn, M. R.,
Yue, H., Warren, B. A., Nguyen, D. B. and Walla, N. K.
TITLE
Protein phosphatases
JOURNAL
Patent: WO 0210363-A 17 07-FEB-2002;
Incyle Genomics, Inc. (US)
FEATURES
source
location/Qualifiers
1..3766
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
/note="Incyle ID No: 7480570CB1"
ORIGIN
Alignment Scores:
Pred. No.: 0 Length: 3766
Score: 472.00 Matches: 663
Percent Similarity: 99.40% Conservative: 0
Best Local Similarity: 99.40% Mismatches: 2
Query Match: 70.98% Indels: 4
DB: 6 Gaps: 0
US-10-029-345A-109 (1-665) x AX374994 (1-3766)
QY 1 MetAlaHsGlnMetLysGlyThrGlnLeuValThrGlnArgLeuValAlaLeuLeuGln 20
DB 538 ATGGCCAGATGAGATGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAG 597
QY 21 SerGlyThrGlnLysValLeuLeuLysSerArgProPheValGlnLysThrSer 40
DB 538 AGTGAAGCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 657
QY 41 HsLysLeuGlnAlaLysAlaGlnLysSerLysLeuMetLysLeuMetLysArgArgLeuGln 60
DB 658 CACATTTTGAAGCCATTAATATCACTGCTCAAGCTTATAGCGAAGAGAGAGAGAGAGAGAG 717
QY 61 AapLysValLeuLysThrGlnLeuLeuLysSerLysLeuMetLysLeuMetLysArgArgLeu 80
DB 718 GACAAAGGTTAATTAACAGAGCTCACTCAAGCTTCAAGCAAGCAATTAAGCTTCACTTGA 777
QY 81 CysSerGlnLysValValValLysArgLysSerGlnAaspValAlaSerLeuSer 100
DB 778 TGCACTGCAAGAGTGTGTGATTAACATCAAGCTCCAAAGATGTTGCCCTCTCTCTTCA 837
QY 101 AapCysPheLeuThrValLeuLeuGlnLysLeuGlnLysSerPheAaspSerValHsLeu 120
DB 838 GACTGTTTCTACGTACTTCTGAGTAACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 897
QY 121 LeuAlaGlyLysPheAlaGlnPheSerArgCysPheProGlyLeuCySerGlnGlyLysSer 140
DB 898 CTTCAGAGGAGGTTTGTGAGAGTCTCTGCTTCTTCCCTGAGCTTGTGAAGAGAGAGAGAG 957
QY 141 ThrLeuValProThrCysLysSerGlnProCysLeuProValAlaAaspLysGlyProThr 160
DB 958 ACTTACAGTCCCTACCTGATTTCTGAGCTTCTGAGCTTCTGAGCTTCTGAGCTTCTGAGCT 1017
QY 161 ArgLysLeuProAaspLeuThrLeuGlnLysGlnAaspValLeuAaspLysLeuLeu 180
DB 1018 CGAATTTCTCCCAATCTTATCTTGGCTGCGCCAGAGAGAGAGAGAGAGAGAGAGAGAG 1076
QY 181 -GlnGlnAaspGlyLysGlyValLeuAaspLysSerThrCysProLysProAasp 200
DB 1077 GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1135
QY 200 HeLysProGlnSerHsPheLeuArgValProValAaspAaspPheCysGlyLysLeu 220
DB 1136 TTATCCCGAGAGCTCATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1195
QY 220 euProThrLeuAaspLysSerValAaspPheLysLeuLysValLysValLysSerHsGlyCy 240

```

```

Db      1196 TGCCGTGGTGGACAAATCAGTAGATTTCATTGAGAAAGCAAAAGCCTCCAAATGAGATGTC 1255
Qy      240  allleuValHisGyleuValIGlyIleSerArgSerAlaThrIleAlaIleTyrIleu 260
Db      1256 TTTCAAGGACCTGTTTAGCTGGAGATCTCCCGCTCCGACCATCGCTATCGCTACATCA 1315
Qy      260  etlySarMetAspMetSerIleuAspGluAlaTyrArgPheValLysGluLysArgPro 280
Db      1316 TGAAGAGATGAGCATCTCTTAGATGAACCTTACAGATTGTGAAAGAAAAGAAAGCTTA 1375
Qy      280  hrIleSerProAsnPheAsnPhelenuGlyIleuLeuLeuAspTyrGlyLysIleLys 300
Db      1376 CTATATCTCCAAACTTCAATTTTCTGGGCAACTCCGACCTGACCTAGAGAAAGATTAA 1435
Qy      300  snGlnThrGlyAlaSerGlyProLysSerIleuLysLeuLeuHisIleuGluLysPro 320
Db      1436 ACCAGACTGAGAGATCAGGGGCCAAAGAGCAAACTCAAGCTGACCTGGAGAAAGCCAA 1495
Qy      320  snGluProValProAlaValSerGluGlyGlyIleuLysSerGluThrProLeuSerPro 340
Db      1496 ATGAACCTGTCCCTGCTGTCTCAGAGGGTGGACAGAAAGCGAGAGCCCTCAGTCCAC 1555
Qy      340  roCyAlaAspSerAlaThrSerGluAlaAlaGlyIleuArgProValHisProAlaSerV 360
Db      1556 CCTGTGCCGACTCTGCTACTCAGAGGACAGAGCAAAAGGCCGCTGACATCCGCCAGCG 1615
Qy      360  alProSerValProSerValGlnProSerIleuLeuLysAspSerProLeuValGlnAla 380
Db      1616 TGCCCAAGCTGCCAGAGTGACCCGCTGTGTAAAGACAGCCCGCTGGATACAGGCC 1675
Qy      380  euSerGlyLeuHisIleuSerAlaAspArgIleuGluAspSerAlaLysIleuLysArgSer 400
Db      1676 TCAGTGGGCTGCACCTGTCCGAGACAGGCTGGAACAGCAATAGCTCAAGCGTTCT 1735
Qy      400  heSerIleuAspIleuLysSerValSerTyrSerAlaSerMetAlaAspIleuHisGlyP 420
Db      1736 TCTCTTGATATCAATCAATCAGTTTCATATTACGCCAGCATGGCAGCATCTTACATGCT 1795
Qy      420  heSerSerSerGluAspAlaLeuGluLysTyrIleuProSerThrThrIleuAspGlyThr 440
Db      1796 TCTCCATCAGAAAGATGCTTTGAAATCTACAAACTTCCACTCTGAGATGGAGACA 1855
Qy      440  snLysLeuGlyGlnPheSerProValGlnGluLeuSerGluGlnThrProGluThrSerP 460
Db      1856 ACAAGCTATGCCAGTCTCCCGCTGTCAAGAACTATCGAGAGACTCCCGAAACAAGTC 1915
Qy      460  roAspLysGluGluAlaSerIleProLysLysLeuGlnThrAlaArgProSerAspSerG 480
Db      1916 CTGATTAAGAGAGAGAGCCAGCATCCCAAGAGCTGCAGACCCGCCAGGCTTCAGACAGC 1975
Qy      480  InsSerLysArgLeuHisSerValArgThrSerSerSerGlyThrAlaGlnArgSerLeu 500
Db      1976 AAGAGCAAGGATTCGATTCGGTCAAGACAGACAGAGTGAGCCCGCCCAAGAGTCCCTT 2035
Qy      500  euSerProLeuHisArgSerGlySerValGluAspAsnTyrHisThrSerPheLeuPheG 520
Db      2036 TATCTCACTGCATCGAAGTGGAGGTGGAAGACAAATTACCAACACAGCTTCCTTTTG 2095
Qy      520  LysLeuSerThrSerGlnGlnHisIleuThrLysSerAlaGlyLeuGlyLeuLysGlyTyrP 540
Db      2096 GCGTTTCCACCGACGACGACCTTCACGAACTGCTGGGCTCTTAAGAGGCTGGC 2155
Qy      540  IsSerAspIleuAlaProGlnThrSerThrProSerLeuThrSerSerThrProPhe 560
Db      2156 ACTCGGATATCTTGGCCCCCGACCTCTTACCCCTTCCTGACCGAGCTGGATATTTTG 2215
Qy      560  IatH-GluSerSerHisPheTyrSerAlaSerAlaIleTyrGlyGlySerAlaSerTyrS 580
Db      2216 CCACAGAGTCTCCACACTTACTCTGCTCAGCATCTTACGAGAGGAGCTGCAGTTACT 2275
Qy      580  eAlaIleTyrSerCySerGlnLeuProThrCyGlyAspGlnValTyrSerValArgArg 600

```

```

Db      2276 CTGCCTACAGCTGCAGCCAGCTGCCCACTTGGGAGACCAAGTCTATTCTGTGGCAGGC 2335
Qy      600  rGlnLysProSerAspArgAlaAspSerArgArgSerTyrHisGluGluSerProPheG 620
Db      2336 GGCAGAGCGCAAGTACAGAGCTGACTCGCGCGAGCTGCATGATGAAGAGAGCCCTTTG 2395
Qy      620  LuLysGluPheLysArgArgSerCyGlnMetGluPheGlyGluSerIleMetSerArgL 640
Db      2396 AAAGCAGATTTAACGCAAGAGCTGCCAAATGAAATTTGGAGAGGCAATGTCAGAGA 2455
Qy      640  snArgSerArgGluGluLeuGlyLysValGlySerGlnSerSerPheSerGlySerMetG 660
Db      2456 ACAGGTCACGGAGAGAGCTGGGGAAAGTGGCAAGTCACTAGCTTTCCGGCAGCATGG 2515
Qy      660  LuIleIleGluValSer 665
Db      2516 AAATCATTTAGAGTCTCC 2532

RESULT 16
BD171157
LOCUS      BD171157              4790 bp    DNA             linear    PAT 17-JAN-2003
DEFINITION Novel gene and protein encoded thereby.
ACCESSION  BD171157
VERSION    BD171157.1  GI:27876969
KEYWORDS   WO 02052005-A/13.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 4790)
            Ohara, O., Nagase, T. and Nakajima, D.
            Novel gene and protein encoded thereby
            Parent: WO 02052005-A 13 04-JUL-2002;
            KATUKA DNA RESEARCH INSTITUTE, OSAMU OHARA, TAKAHIRO NAGASE, DAISUKE
            NAKAJIMA
COMMENT     OS Homo sapiens (human)
            PN WO 02052005-A/13
            PD 04-JUL-2002
            PF 20-DEC-2001 WO 2001JP011217
            PR 22-DEC-2000 JP 00P 389742
            PI OSAMU OHARA, TAKAHIRO NAGASE, DAISUKE NAKAJIMA
            PC C12N15/12, C07K14/47
            CC Novel gene and protein encoded thereby
            FH Key location/Qualifiers
            FT CDS location/Qualifiers
               (184).. (2178).

FEATURES
            source          1..4790
                           /organism="Homo sapiens"
                           /mol_type="genomic DNA"
                           /db_xref="taxon:9606"

ORIGIN
Alignment Scores:
Pred. No.:      0
Score:          472.00
Percent Similarity: 99.40%
Best Local Similarity: 99.40%
Query Match:    70.98%
DB:             6
Gaps:          0

US-10-029-345A-109 (1-665) x BD171157 (1-4790)
Qy      1  MetAlaHisGluMetIleGlyThrGlnIleValThrGluArgLeuValAlaLeuLeuGlu 20
Db      184  ATGGCCCATAGATGATGGAACTCAAAATTTGTTACTGAGAGGTTGGCTGCTGCTGGAA 243
Qy      21  SerGlyThrGluLysValLeuLeuLysSerArgProPheValGluTyrAsnThrSer 40
Db      244  AGTGAAGCAAGAAAAGTCTGCTATTTGATAGCCGCGCATTTGTGAAATACATATCATCC 303
Qy      41  HisIleuGluAlaIleAsnIleAsnCySerLysLeuMetLysArgArgLeuGlnGln 60
Db      304  CACATTTGGAAAGCCATTATATCACTGCTCCAAAGCTTATGAAGCGAAGTTGCAAG 363

```


[illegible]

QY	420	heserSerSeGluuAaPrAlaLeuGluYrTyrlYrPProSerThThThLeuAspGlyThra	440
Db	1442	TCTCTCTCATCGAAGAATGCTTTGGAAATACTAACAACTTTCACATCTTGATGGAGACA	1501
QY	440	anLYbLeuCYGglinPhSerSerProValGngluLeuSerGluGlnThProGluThSerP	460
Db	1502	ACAAAGCTATGGCAAGTTCTCCCCCTGTTCAAGAACTATGGAGCAAGCTCCGAAACCACTC	1561
QY	460	roAspLYgGluGluAlaSerllLeProLYbLYLeuGlnThAlaArgProSerApsSerG	480
Db	1562	CTGATTAAGAGAAAGAACCGACATCCCAAGAAAGCTGACAGACCGCAGGCTTCACAGACCC	1621
QY	480	InserLYbArgLeuHHisSerValArgThSerSerSerGlyThAlaGlnArgSerLeuL	500
Db	1622	AGAGCAAGCGATTCATTCGCTGCAAGAACACAGACAGTGGCACCGCCAGAGTCCCTTT	1681
QY	500	euserProLeuHHisArgSerGlybSerValGluAaPaanYrHIsThSerPheLeuPheG	520
Db	1662	TATCTCCACGTCACTGAAGTGGAGCGGTGGAGGCAATPACACACCAAGCTTCCTTTTCG	1741
QY	520	LYLeuSerThSerGlnGlnHIsLeuThLYrLYbSerAlaGlyLeuGlyLYLeuLYgLYTrpH	540
Db	1742	GCTTTTCCACAGCCAGCAGAGACACTCAGAAAGTCTGCTGGCTGGGCTTAAGGGCTGCG	1801
QY	540	IsSerAspLYLeuAlaAProGlnThSerThrProSerLeuThSerSerTrpTYrPhea	560
Db	1802	ACTCGSATATCTTGAGCCCCCAGACCTCTACCCCTTCCTGACCAGCAGCTGTATTTTG	1861
QY	560	laThrGluSerSerHIsPheTYrSerAlaSerAlaIleTYrGlyLYSerLYAserTYrs	580
Db	1862	CCACAGAGTCTCTCACACTTCTACTCTGCTCAGCATCTACGAGGACAGTGGCCAGTTACT	1921
QY	580	erAlaTYrSerCYbSerSeGluLeuPProThrCYgLYAspGlnValTYrSerValArgArg	600
Db	1922	CTGCCCTACAGCTGCAGCCAGCTGCCACTTGCCGAGACCAAGTCTATTTCTGGCGCAGC	1981
QY	600	rgGlnLYbProSerAspArgAlaAspSerArgArgSerTrpHIsGluGluSerProPheG	620
Db	1982	GGCAGAGCCCAAGTGACAGAGCTGACCTCGCGCGGAGCTGGCATGAGAGAGCCCTTTG	2041
QY	620	LYbLYglinPheLYbArgArgSerCYbGlnMetGluPheGlyGluSerllLeuMetSerGluA	640
Db	2042	AAAAGCAGATTTAAACGACAGAGCTGCCAAATGAAATTTGGAGAGAGCATCTATGTCAGAGA	2101
QY	640	anArgSerArgGluGluLeuGlyLYbValGlySerGlnSerSerPheSerGlySerMetG	660
Db	2102	ACAGGTCCGGGAAAGAGCTGGGGAAAGTGGCACTCACTTTCGCGGACAGATGG	2161
QY	660	LYbLeuLeuGluValSer	665
Db	2162	AAATCATTTGAGGTCTCC	2178
RESULT 17			
BD183422			
LOCUS			
DEFINITION Novel genes and proteins encoded by the genes.			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
OS Homo sapiens (human)			
PN JP 2002345492-A/135			
PD 03-DEC-2002			
PI 26-FEB-2002 JP 2002049009			
PI OSAMU OHARA, TAKAHIRO NAGASE, DAISUKE NAKAJIMA			

PC C12N15/09, C07K14/47//A61K31/711, A61K38/00, A61K48/00, A61P25/00,
 PC A61P25/14,
 PC A61P25/18, A61P35/00, C12N15/00, A61K37/02
 CC Novel genes and proteins encoded by the genes FH Key
 Location/Qualifiers
 FT CDS (184)..(2178).

FEATURES
 source 1..4790
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 4790
 Score: 472.00 Matches: 663
 Percent Similarity: 99.40% Conservative: 0
 Best Local Similarity: 99.40% Mismatches: 2
 Query Match: 70.98% Indels: 4
 DB: 6 Gaps: 0

US-10-029-345a-109 (1-665) x BD183422 (1-4790)

QY 1 MetAlaHisGluMetIleGlyThrGlnIleValThrGluArgLeuValAlaLeuLeuGlu 20
 Db 184 ATGCCCATGATGATGATTGGAACTCAAAATTGTTACTGAGAGGTGGCTGCTGCTGAA 243
 QY 21 SerGlyThrGluSerValLeuLeuIleAspSerArgProPheValGluTyrAsnThrSer 40
 Db 244 AGTGAACCGAAGAAAGTGTGCTAATGTAAGCCGCAATTTGTGAATAACATACATCC 303
 QY 41 HisIleLeuGluAlaIleAsnIleAsnCysSerIleLeuMetLysArgIleGlnGln 60
 Db 304 CACATTTTGGAAACCATTAATCAACTGCTCCAACTTATGAGCGAAGTTGCAACG 363
 QY 61 AspIleValIleuIleThrGluLeuIleGlnHisSerAlaIleHisIleValAspIleAsp 80
 Db 364 GACAAAGTGTATATACAGAGCTCATCCAGCATTCAGCGAAACATAGGTTGCAATTGAT 423
 QY 81 CysSerGlnIleValIleValIleValIleValIleValIleValIleValIleValIleVal 100
 Db 424 TGCAGTCCAGAAAGGTGTAGTTACGATCAAGAGTCCCAAGATGTGCTCTCTCTTCA 483
 QY 101 AspCysPheLeuThrValLeuLeuGluGlyIleLeuGluIleLysSerPheAsnSerValHisIleu 120
 Db 484 GACTGTTTCTCACTGTACTCTCGGTTAACTGAGAAAGCTTCAACTCTGTTCACTG 543
 QY 121 LeuAlaGlyGlyPheAlaGluPheSerArgCysPheProGlyLeuCysGluGlyLysSer 140
 Db 544 CTTCAGGTGGGTGCTGAGTCTCTCGTGTTCCTGCGCTCTGTGAAGAAATCC 603
 QY 141 ThrIleValIleProThrCysIleSerGlnProCysIleProValAlaAsnIleGlyProThr 160
 Db 604 ACTCTAGTCCCTACCTGCAATTTCTCGCTTCTTAACCTGTTGCCAACAATGGGCAACC 663
 QY 161 ArgIleLeuProAsnLeuTyrLeuGluCysGlnArgAspValLeuAsnLysGluLeuIle 180
 Db 664 CCAATCTTCCCAATCTTATCTTGGCTGCCAGCAGAGTCTCTCAACAGAGAGCTGAT- 722
 QY 181 -GlnGlnGlnGlyIleGlyTyrValIleuAsnAlaSerTyr-ThrCysProLysProAsp 200
 Db 723 GGAGCAGATGGATGGATTATGTATGAATGCCAGAA-TACCTGTCCAAAGCTGACT 781
 QY 200 HeIleProGluSerHisPheLeuArgValProValAsnAspSerPheCysGluLysIleIle 220
 Db 782 TTAATCCCGAGTCTCAATTTCTGCGGTGCTGTAATGACGCTTTGTGAGAAATTT 841
 QY 220 euProTyrLeuAspLysSerValAspPheIleGluLysAlaLysAlaSerAsnGlyCysV 240
 Db 842 TGCCGTGGTGGCAAACTAGTATTTCTTGAAGAAAGCAAAAGCTCCCAATGATGTG 901
 QY 240 alLeuValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIleW 260

Db 902 TTCTAGTGCACCTGTTTAGCTGGGATCTCCCGCTCCGCCACCATTCGCTTACATCA 961
 QY 260 eLysArgMetAspMetSerLeuAspGluAlaTyrArgPheValLysGluLysArgProT 280
 Db 962 TGAAGAGATGACATGCTCTTATGATGAAGCTTACAGATTGTGTAAGAAAGAAAGCACTTA 1021
 QY 280 HisIleSerProAsnPheAsnPheLeuGluGlnIleLeuAspTyrCylLysLysIleLysA 300
 Db 1022 CTATATCTCCAAACTTCAATTTTCTGGGCGCAACTCTGACCTATGAGAAAGATTAAGA 1081
 QY 300 enGlnThrGlyAlaSerGlyProLysSerLysIleLysLeuLeuHisIleuGluLysProA 320
 Db 1082 ACCAAGCTGAGACATCAAGGCCCAAGACAACTCAACTGCTGACCTGAGAAAGCCAA 1141
 QY 320 enGluProValProAlaValSerGluGlyGlnLysSerGluThrProLeuSerProp 340
 Db 1142 ATGAACCTGCTCCCTGCTCTGAGAGGTGAGCAAGAAAGCGAGCGCCCTCAGTCCAC 1201
 QY 340 roCysAlaAspSerLysThrSerGluAlaIleGlyGlnLysSerGluThrProLeuSerProp 360
 Db 1202 CTGTGCGACCTGCTACCTCAAGAGCGAGCAAGAAAGCCCGTGCATCCCGCAAGCG 1261
 QY 360 alProSerValProSerValGlnProSerLeuLeuGluAspSerProLeuValGlnAlaL 380
 Db 1262 TGCCAGGAGTCCCAAGCGTGCAGCGCTGCTGTTAGAGACAGCCCGCTGTTACAGGCGC 1321
 QY 380 euSerGlyLeuHisIleuSerAlaAspArgLeuGluAspSerAsnLysLeuLysArgSerP 400
 Db 1322 TCAGTGGGCTCACCCTGCTCCAGAGCGCTGGAAGACAGCAATAGCTCAAGCGTTCT 1381
 QY 400 heserLeuAspIleLysSerValSerTyrSerAlaSerMetAlaIleSerLeuHisIleGlyP 420
 Db 1382 TCTCTCTGATATCAATCACTTCAATATTCAGCAGCATGGCAGATCTTACATGGCT 1441
 QY 420 heserSerSerGluAspAlaLeuGluTyrTyrLysProSerThrThrLeuAspGlyThra 440
 Db 1442 TCTCTCATCAGAAAGTCTTGGAAATCAACAACTTCCACTGATCTGATGGACCA 1501
 QY 440 enLysLeuCysGlnPheSerProValGlnLysLeuSerGluGlnThrProGluThrSerP 460
 Db 1502 ACAAGCTATGCAAGTCTCCCTGTTCAAGAACTATGAGAGCAACCTCCGAAACCAAGTC 1561
 QY 460 roAspLysGluAlaIleSerIleProLysLysLeuGlnThrAlaArgProSerAspSerG 480
 Db 1562 CTGATTAAGAGAGAACCAAGCATCCCAAGAAAGCTGCAAGCCGAGGCTTTCAGACAGCC 1621
 QY 480 InsertlySargLeuHisSerValArgThrSerSerSerGlyThrAlaGlnArgSerLeuL 500
 Db 1622 AGACCAAGCGATTCATTCGCTGAGAACCAAGCAAGCAAGTGGACCGCCAGAGTCCCTTT 1681
 QY 500 euSerProLeuHisArgSerGlySerValGluAspAsnTyrHisThrSerPheLeuPheG 520
 Db 1682 TATCTCACTGCAATGAGAGGTGGAGCGTGAAGCAATTCACACACAGCTTCTCTTTTCG 1741
 QY 520 LysLeuSerThrSerGlnGlnHisIleuThrLysSerAlaGlyLeuGluLysLeuGlyTyrP 540
 Db 1742 GCTTTTCAACAGCAGCAGACACTTCAAGAGTCTGCTGGGCTTAAAGGCTGGG 1801
 QY 540 isSerAspIleLeuAlaProGlnThrSerThrProSerLeuThrSerSerTyrPheA 560
 Db 1802 ACTCGATATCTTGGCCCCCAAGCCTTACCCCTTCCGACAGCAGAGCTGGATATTTTG 1861
 QY 560 latThrGluSerSerHisPheTyrSerLysAlaSerAlaIleTyrGlyLysSerAlaSerTys 580
 Db 1862 CCAAGAGTCTCTACACTTCACTGCTCAGCAATCTAAGAGGAGGAGGAGGAGGAGGAGTACT 1921
 QY 580 exAlaTyrSerCysSerGlnLeuProThrCysGlyAspGlnValTyrSerValArgArg 600
 Db 1922 CTGCTTCAAGCTGAGCAGCAGCTGCTGCTGCGAGACCAAGTATTTCTGTGGAGGAGC 1981
 QY 600 rgGlnLysProSerAspArgAlaAspSerArgArgSerTyrHisGluGluSerProPheG 620
 Db 1982 GGCAGAGCCAGTGAAGAGCTGACTCGCGGAGGAGCTGGCATGAAGAGGCCCTTTG 2041

QY 620 IuYegInPheLySArGArSeCySeGImeGluPheGlyGluSerIleMeSeGluA 640
| | | | |
Db 2042 AAAAGCATTTAAACCCAGAGCTGCCAAATGGAAATTTGGAGACATCATGTCAGAA 2101
| | | | |
QY 640 snArSeArGrgIuGluLeuGlyLySeValGlySerInSeSerPheSeGlySeMeG 660
| | | | |
Db 2102 ACAGGTCACGGAGAAAGCTGGGAAAGTGGCAGCTCAGTCTACCTTTGGCGCAGCATGG 2161
| | | | |
QY 660 IuIleIleGluValSer 665
| | | | |
Db 2162 AATCATTTGAGTCTCC 2178
| | | | |
RESULT 18
AB051487 4790 bp mRNA linear PRI 07-FEB-2001
LOCUS AB051487
DEFINITION Homo sapiens mRNA for KIAA1700 protein, partial cds.
ACCESSION AB051487
VERSION AB051487.1 GI:12697944
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Nagase, T., Kikuno, R., Hattori, A., Kondo, Y., Okumura, K. and Ohara, O.
TITLE 1 (sites)
Nagase, T., Kikuno, R., Hattori, A., Kondo, Y., Okumura, K. and Ohara, O.
Prediction of the coding sequences of unidentified human genes.
XIX. The complete sequences of 100 new cDNA clones from brain which
code for large proteins in vitro
DNA Res. 7 (6), 347-355 (2000)
JOURNAL MEDLINE 21082932
PUBMED 11214970
REFERENCE
AUTHORS 2 (bases 1 to 4790)
Ohara, O., Nagase, T. and Kikuno, R.
Direct Submision
Submitted (22-NOV-2000) Otsamu Ohara, Kazusa DNA Research Institute,
Department of Human Gene Research; 1532-3, Yana, Kisarazu, Chiba
292-0812, Japan (E-mail:cdnainfo@kazusa.or.jp,
URL: http://www.kazusa.or.jp/huge, Tel:81-438-52-3913,
Fax:81-438-52-3914)
FEATURES
source Location/Qualifiers
1..4790
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="fj15353"
/note="vector:pb1uescriptII SK plus"
1..4790
/gene="KIAA1700"
<109..2181
/gene="KIAA1700"
/note="Start codon is not identified."
/product="KIAA1700 protein"
/protein_id="BAB21791.1"
/db_xref="GI:12697945"
/translation="AEEHTTITSSWQIKGGGKGLIVMAHEMIGTQIVTERVALL
ESGTEKVLIDSPFVEYNTSHLEAININCSKLMRRLOQOKVLTTELIOHSAKRV
DIDCSQVVVYDOSQDVAISLSDCFLTVLGLGKESFNSVHLGAFERCFEGL
CEGKSTLVPTCSIQPCIPANIGPRLILNLYGCGRDVNLKEMQNGIIVLVANAS
TECKPDPFIESHLRVVNDSPCEKILPMLDKSDVPIEKAKSNGCVLHACLAGSRS
ATTALAIKMRDMSLDEAVRYPKERPTISPFPNLGQLLDYEKKIKNOTASGSKS
KKLHLLEKNEPVPVAVSGGQKSETPPLSPCADSATSEAGORPVPVAVSVSPSVQ
PSLDESPVLQALSHLADRLSDNKLKRSFSLDKSVYSASVAASLHGFSSVD
ALEYKPSSTLLDGTNKLQFSPVQELSEGTPEPSPDEKASIPKLTQTPRSDGSKR
LHSVARTSSSGTASLSPSLHRSQSVEDVYHTFLGLSTQOHLTKSAGLKGWS
DLAPOTSPLSTLSWYPATESHPYASAIYGGASVAYSASOLPTCGDQVYSVR
ROKSPRADSRBSRMHESPPEKQPKRSCOMERGESIMENSREELGKVGQSSTSG
SMETIEVS"

ORIGIN
Alignment Scores:
Pred. No.:

0 Length: 4790

Score: 472.00 Matches: 663
Percent Similarity: 99.40% Conservative: 0
Best Local Similarity: 99.40% Mismatches: 2
Query Match: 70.98% Indels: 4
DB: 9 Gaps: 0
US-10-029-345A-109 (1-665) x AB051487 (1-4790)
QY 1 MetAlaHlaGluMeIleGlyThrGluIleValIThrGluArgLeuValAlaLeuLeuGlu 20
| | | | |
Db 184 ATGGCCCATGAGATATGGAATCTCAAAATGTTACTGAGAGAGGTGGTGGCTCTGCTGGA 243
| | | | |
QY 21 SerGlyThrGluValLeuLeuLeuIleApsSerArpProPheValGluTyraenThrSer 40
| | | | |
Db 244 AGTGACGAGAAAAGTGTGCTTAATGATACCGGCCATTGTGTGAATACATATCATC 303
| | | | |
QY 41 HisIleLeuGluAlaIleAenIleAenCySeSerIyLeuMeLyArgrhLeuGlnGln 60
| | | | |
Db 304 CACATTTGGAGCATTATATCAATGCTCCAGCTTATGAGGAAAGTTGCAACAG 363
| | | | |
QY 61 AspLySValIleIleThrGluLeuIleGlnHisSerAlaLyHisLySValAspIleAap 80
| | | | |
Db 364 GACAAAGTGTATTAACAGAGCTCATCCAGCATTCAGCGAAACATNAGGTTGACATGTAT 423
| | | | |
QY 81 CySeSerGlnLySValValValTyraSpGlnSerSerGlnAspValAlaSerIleuSer 100
| | | | |
Db 424 TGCAGTCAGAGAGTTGTATTCATGATCAAACTCCCAAGATGTTGCTCTCTCTTCA 483
| | | | |
QY 101 AspCySphLeuThrValLeuLeuGlyLyLeuGluLySLeuPheAenSerValHisLeu 120
| | | | |
Db 484 GACTGTTTCTCAGTACTCTGAGGTAACTGGAAGAGAGCTTCAACTCTGTCACTG 543
| | | | |
QY 121 LeuAlaGlyLyPheAlaGluPheSerArpCySphProGlyLeuCySeGluLyLySer 140
| | | | |
Db 544 CTTCGAGGAGGAGTTGTCTGAGTCTCTGTTTCCCGGCTCTGTAGAGAAATCC 603
| | | | |
QY 141 ThrLeuValProThrCySileSerGlnProCySLeuProValAlaAenIleGlyProThr 160
| | | | |
Db 604 ACTCTAGTCCCTACCTGCTGCTTCTCAGCTTCTTACCTGTTGCCAACATTGGGCCAAC 663
| | | | |
QY 161 ArgIleLeuProAenLeuTyreGlyCySglnArgAspValLeuAenLySLeuIle 180
| | | | |
Db 664 CCAATTCCTCCCAATCTTATCTTGTGCTGCCAGCGAGATGCTCTCAACAGAGGCTGAT- 722
| | | | |
QY 181 -GlnGlnGlyLyIleGlyTyraValLeuAenAlaSerTyra-ThrCySProLyProAap 200
| | | | |
Db 723 CGAGCAGATGGAGATTGTTATGTATGTTAATCCAGCAA-TACCTGTCCAAAGCCGACT 781
| | | | |
QY 200 heIlePProGluSerHisPheLeuArgValProValAsnApsSerPheCySgLyLyIle 220
| | | | |
Db 782 TTATCCCGAGTCTCATTTCTGCGGTGCTGTGATGACAGCTTTGTGAGAAATTT 841
| | | | |
QY 220 euProTirPheuApySerValaPheIleGluLySAlaLySAlaSerAenGlyCySv 240
| | | | |
Db 842 TCCCGGTGTGACAAATCAGTATGATTTGATGAGAAAGCAAAAGCTCCATATGATG 901
| | | | |
QY 240 alleuValHisCySleuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrlIle 260
| | | | |
Db 902 TTCCTAGTGACAGTTTATGCTGGGATCTCCGCTCCGCCACATGCTATCGCTCATCA 961
| | | | |
QY 260 eLySAlaArgSerApsPheSerIleuAapGluAlaTyraPhePheAlaLySgLyLySArgP 280
| | | | |
Db 962 TGAAGAGATGAGCATGTTTATGATGAGAGCTTACGATTTTGAAGAAAGAAAGCTTA 1021
| | | | |
QY 280 hrIleSerProAenPheAenPheLeuGlyGlnLeuLeuApyTyrgLyLySlyIleY 300
| | | | |
Db 1022 CATATCTCCAAATCTTCAATTTTCTGGGCAACTCTGGAATATGAGAAAGATTAAGA 1081
| | | | |
QY 300 enGlnThrGlyAlaSerGlyProLySLeuLySLeuLeuHisIleGluLyLySProA 320
| | | | |
Db 1082 ACCAGCTGAGACATCAGGCGCAAAAGAGCAAACTCAAGTCTGTGCACTCGAGAGCCAA 1141
| | | | |
QY 320 enGluProValProAlaValSerGluGlyGlyLySLeuSerGluThrProLeuSerPro 340
| | | | |

```

Db      1142 ATGAACCTGTCCTGCTGCTCAGAGGCTGACAGAAAACGAGACGCCCTCACTGCTCAC 1201
QY      340  rocyaaiaaspseraiaathsergiuaiaaiaaglyinaaprovahisproalaser 360
Db      1202 CCGTGCCGACTGCTGCTACTCAGAGAGCAGACAGAACAGCCCGCTGCACTCCGCCACGCG 1261
QY      360  alproserValproserValglproserleuengluaapserProleuValglua 380
Db      1262 TGCCACAGCGTCCAGAGGTCCAGCCGCTGCTTAGAGACACCCGCTGTGACAGGCG 1321
QY      380  euSerGlyleuHisleuSerAlaaspArgleuGluaapseranlyleuLyAspArgser 400
Db      1322 TCAGTGGCTGCACTGCTCCGACAGAGCTGGAAACAGCAATAGCTCAAGCGTTCCCT 1381
QY      400  heserleuaspilleysserValSerTySerAlaserMetAlaaserleuHisGlyP 420
Db      1382 TCTCTCGATATCAATCAATCAGTTTCATTCACCCAGCATGGACATCTTACATGGCT 1441
QY      420  heserSerSerGluaspAlaLeuGlulYrTyTySerProserThrThreAspGlyThra 440
Db      1442 TCTCCATCAGAAAGATGCTTTGGATATCTACAAACCTTCACTCTGATGGACCA 1501
QY      440  smlyleuCySglnPheSerProValGlnGluLeuSerGluGlnThrProGluThrSerP 460
Db      1502 ACAAGCTATGCCAGTTCTCCCTGTTACAGAACTATCGAGACAGACTCCCGAAACAGTC 1561
QY      460  roAspLySglnGluAlaserIleProLySglnleuGlnThrAlaArgProserAspSerG 480
Db      1562 CTGATTAAGAGAGAGCAGATCCCAAGAGAGCTGACAGCCGACAGGCTTACAGACAGCC 1621
QY      480  InSerLySglnleuHisSerValArgThrSerSerSerGlyThrAlaGlnArgSerLeu 500
Db      1622 AGAGCAAGCGATTCATTCGGTCAGAAACAGACAGAGCTGGACCGCCAGAGCTCCCTT 1681
QY      500  euSerProleuHisAspSerGlySerValGluAspAsnTyRhisThrSerPheleuPheG 520
Db      1682 TATCTCCTACCTGATCCAGATGGAGGCTGAGAGACATTAACACACMACCTTCTTTTCG 1741
QY      520  YleuSerThrSerGlnGlnHisleuThrLySglnSerAlaGlyleuGlyleuLySglnThrP 540
Db      1742 GCTTTCCACACAGCCAGCAGCACCTTCAGAACTGCTGGCCCTTAAGGCGCTGGC 1801
QY      540  iSerAspIleleuAlaProGlnThrSerThrProserleuThrSerSerTrpTyPhea 560
Db      1802 ACTCGGATATCTGGCCCCCAGACCTTACCCTTCCCTGACACAGAGCTGGATTTTG 1861
QY      560  laThrGluSerSerHisPheTyRSerAlaSerAlaIleTyRGlYglYSerAlaSerTyRS 580
Db      1862 CCACAGAGTCTTCACACTTCTACTCTGCTCAGCCATCTACGAGAGCAGTGCACATTACT 1921
QY      580  eAlaTySerCySerGlnleuProThrCySglnAspGlnValTyRSerValaArga 600
Db      1922 CTGGCTACAGCTGACCCAGCCAGCTGCCCTTGGCGAGACCAAGTCTATTCCTGGCCAGCC 1981
QY      600  xGlnLyAspProserAspArgAlaaspSerArgArgSerTrpHisgluGluSerProPheG 620
Db      1982 GGCAAGAGCCAGAGTACAGAGCTGACTGCGCGGAGCTGGCATGAAGAGAGCCCTTTTG 2041
QY      620  JulYglnPheLyArgArgSerCySglnMetGlnPheGlyGluSerIleMetSerGluA 640
Db      2042 AAAACAGATTAAACGACAGAACTCCCAATGTGAATTTGGAGAGCATCAATGTCAGAGA 2101
QY      640  snArgSerArgGluGluLeuGlyLySValGlySerGlnSerSerPheSerGlySerMetG 660
Db      2102 ACAAGTCAAGGAGAAAGCTGGGGAAGTGGGAGTCACTAGCTTTTGGGCAACATGG 2161
QY      660  JulIleIleGluValSer 665
Db      2162 AAATCATTAGAGTCTCC 2178

```

RESULT 19
AX099933

```

LOCUS      AX099933                2118 bp    DNA                linear    PAT 02-APR-2001
DEFINITION Sequence 15 from Patent WO0120004.
ACCESSION  AX099933
VERSION     AX099933.1  GI:13538943
KEYWORDS   SOURCE
ORGANISM   Homo sapiens (human)
REFERENCE  1
AUTHORS    Yue,H., Tang,Y.T., Bandman,O., Hillman,J.L., Baughn,M.R.,
            Azimzai,Y. and Lu,D.A.
TITLE       Protein phosphatase and kinase proteins
JOURNAL     Patent: WO 0120004-A 15 22-MAR-2001;
            Incyte Genomics, Inc. (US)
FEATURES   source
            1..2118
            /organism="Homo sapiens"
            /mol_type="unassigned DNA"
            /db_xref="taxon:9606"
            /note="Incyte ID NO: 1234795CBI"

ORIGIN
Alignment Scores:
Pred. No.:      0          Length:      2118
Score:          439.00    Matches:      663
Percent Similarity:  98.96%  Conservative:  0
Best Local Similarity:  98.96%  Mismatches:    2
Query Match:      66.02%      Indels:      7
DB:               Gaps:      0

US-10-029-345A-109 (1-665) x AX099933 (1-2118)
QY      1  MetAlaHisGluMetIleGlyThrGlnIleValThrGluArgLeuValAlaLeuGlu 20
Db      25  ATGGCCCAATGAGATGATGGAACTCAAAATGTTTCTAGAGAGTTGGCTGCTGCGAA 84
QY      21  SerGlyThrGluLySValleuLeuIleAspSerArgProPheValGluTyRAsnThrSer 40
Db      85  AGTGGACGAAAGAAAGTCCCTTAATGATAGCCGACATTTGGTAATCAATACATACC 144
QY      41  HisIleleuGluAlaIleAsnIleAsnCySerLySleuMetLySArgArgleuGln 60
Db      145  CACATTTTGGAAAGCCATTAATATCACTGCTCCAGCTTATGAAGCAAGGTTGCCAACAG 204
QY      61  AspLySValleuIleThrGluLeuIleGlnHisSerAlaLySHisLySValAspIleAsp 80
Db      205  GACAAAGTTTAATTACAGAGCTCATCAGCATTCACAGCAACATTAAGTTGACATTGAT 264
QY      81  CySerGlnLySValValIleTyRAspGlnSerSerGlnAspValAlaSerLeuSerSer 100
Db      265  TGCAGTCAGAGAGTTGATGATTACGATCAAAAGCTCCCAAGATGTGCTCTCTCTTCA 324
QY      101  AspCySphleuThrValleuLeuGlyLySleuGlyLySleuSerPheAsnSerValHisleu 120
Db      325  GACTGTTTTCTCACTGATCTTGGGTAACCTGGAGAAAGCTTCAACTGTGTCACTG 384
QY      121  LeuAlaGlyIlePheAlaGluPheSerArgCySphProGlyLeuCySglnGlyLySser 140
Db      385  CTGGCAGGTGGGTTGTGTGATCTCTCGTTGTTTCCCTGCTGTCGTGAAGAAAAATCC 444
QY      141  ThrleuValProThrCySHisSerGlnProCyLeuProValAlaAsnIleGlyProThr 160
Db      445  ACTCTAGTCCCTTACTGATCTTCTCAGCTTGTCTTACTTGTCCCAATGGGCCAAC 504
QY      161  ArgIleuProAsnleuTyRleuGlyCySglnArgAspValleuAsnLySglnleuIle 180
Db      505  CGAATCTTCCCAATCTTATCTTGGCTGCCACGAGATGCTCTCAACAAGAGCTGAT- 563
QY      181  -GlnGlnAsnGlyIleGlyTyRValleuAsnAlaSerTyR-ThrCySProLySProAsp 200
Db      564  GCAGCAGAAATGGAGATTGGTTATGTGTTAAATGCACAA-TACTCTGTCCAAACCTGACT 622

```

QY 200 he11eProgluSerH1sPheLeuArgValProValAsnAspSerPheCysGluValLeu 220
 DB 623 TTTATCCCCGAGTTCATTTCTCGTGTGCTTGAAATGACACTTTTGTGACAAAATTT 682
 QY 220 euProTrpLeuAspLysSerValAspPhe11eGluValAlaValAsnAsnGlyCys 240
 DB 683 TGGCGGTGGTTGACCAATCATGTGATTTCAATTGAAAGCAAAAGCTCCAAATGATGG 742
 QY 240 a11eValH1sCysLeuAlaGly11eSerArgSerAlaThr11eAla11eAlaTyr11e 260
 DB 743 TTTCTATGACATGTTTAACTGGGATCTCCGCTCCGACCATCGCTATCGCTTACATCA 802
 QY 260 e11yAspArgMetAspMetSerLeuAspGluAlaTyrArgPheVal11yGluValArgPro 280
 DB 803 TGAAGAGATGACATGTCTTTAGATGAACTTACAGATTTGTGAAGAAAAGACCTA 862
 QY 280 hr11eSerProAsnPheAsnPheLeuGlyGluLeuLeuAspTyrGluVal11yAla 300
 DB 863 CTATATCTCCAACTTCAATTTTCTGGGCCAATCTCGAGCATGAGAAAGATTAAAG 922
 QY 300 enGlnThrGlyAlaSerGlyProLysSerLysLeuLysLeuH1sLeuGluLysPro 320
 DB 923 ACCAGACTGGAGCATCAGGGCCAAAGACCAAGCTGAGCTGACCTGGAGAAAGCCAA 982
 QY 320 enGluProValProAlaValSerGluGlyGluLysSerGluThrProLeuSerPro 340
 DB 983 ATAAACCTGCTCCTGCTGCTCAGAGGGTGGACAAAGAAAGCGAGAGCCCTCAGTCCAC 1042
 QY 340 roCysAlaAspSerAlaThrSerGluAla11eGlyGlnArgProValH1sProAlaSer 360
 DB 1043 CCGTGCCGACCTGTGCTACCTCAGAGGACAGACAAAGCCCGTGACCTCCGCCACAG 1102
 QY 360 a1ProSerValProSerValGlnProSerLeuGluLysAspProLeuValGlnAla 380
 DB 1103 TGCCTCAGCGTCCAGGCTGACACCGCTGTAAAGAGCAAGCCGCTGTGTCAGAGCG 1162
 QY 380 euSerGlyLeuH1sLeuSerAlaAspArgLeuLysAspSerLeuLysValArgSer 400
 DB 1163 TCGTGGGCTGACCTGTCCGACAGAGGCTGGAAAGCAATTAAGCTTAAGCTTCCCT 1222
 QY 400 heSerLeuAspLysLeuSerValSerTyrSerAlaSerMetAlaAlaSerLeuH1sGly 420
 DB 1223 TCTCTGTGATATCAATCATGTTTCAATTCACAGCAGCAGCATCTTACATGCT 1282
 QY 420 heSerSerSerGluAspAlaLeuGluLysTyrLysProSerThrThrLeuAspGlyTh 440
 DB 1283 TCTCCTCATAGAAAGATGCTTGGAAATACAAACCTTCCACTACTGAGTGGAGCA 1342
 QY 440 enLysLeuCysGlnPheSerProValGlnGluLeuSerGluGlnThrProGluThrSer 459
 DB 1343 ACAAGCTATGCGAGTCTCCCTGTTCAAGAACTATCGAGACAGACTCCGAAACCAAGT 1402
 QY 460 -ProAspLysGluGluAlaSer11eProLysLysLeuGlnThrAlaArgProSerAsp 479
 DB 1403 CCTGATTAAGAGAGAGCCAGCATCCCAAGAAAGCTGACAGCCAGGCTTCAACAG 1462
 QY 479 rgLysLysValArgLeuH1sSerValArgThrSerSerSerGlyThrAlaGlnArgSer 499
 DB 1463 CCGAGACCAAGCGATTCATTCGCTCAGAACCAAGACAGAGTGGCCAGGAGTCCCT 1522
 QY 499 u1euSerProLeuH1sArgSerGlySerValGluAspAsnTyrH1sThrSerPheLeu 519
 DB 1523 TTTATCTCCATCTCATGAGAGTGGAGCGTGGAGCAATTAACCAAGCTTCTCTTT 1582
 QY 519 eGlyLeuSerThrSerGlnGlnH1sLeuThrLysSerAlaGlyLeuGlyLysLeuGly 539
 DB 1583 CGGCTTTTCCACCAAGCAGACAGACATCAAGAGTCTGCTGGCCCTGGCCCTTAAAGCT 1642
 QY 539 rph1sSerAspLysLeuAlaProGlnThrSerThrProSerLeuThrSerSerTyrTyr 559
 DB 1643 GGCAGCTCGATATCTTGGCCCCCAGACCTTACCCCTTCCAGACAGAGCTGATATT 1702
 QY 559 he1a1aThrGluSerSerH1sPheTyrSerAlaSerAla11eTyrGlyGlySerAlaSer 579

DB 1703 TTGCCACAGAGCTCTACACTTCTACTCTGCTGACCATCTTACAGAGGACAGGCCAGT 1762
 QY 579 yrSerAlaTyrSerCysSerGlnLeuProThrCysGlyAspGlnValTyrSerValArg 599
 DB 1763 ACTTGCTCTACAGCTGACAGCCAGCTGCCACTTGGCGAAGCAAGCTTCTTGTGTCGCA 1822
 QY 599 rGArgGluLysProSerAspArgAlaAspSerArgArgSerTrpH1sGluGluSerPro 619
 DB 1823 GCGCGCAGAGCCAGTGAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 1882
 QY 619 heGluLysGlnPheLysArgArgSerCysGlnMetGluPheGluSer11eMetSer 639
 DB 1883 TTGAAGAGCATTTTAAAGCAGACAGCTGCAATGGAATTTGGAGAGACATCATCTCAG 1942
 QY 639 lAsnArgSerArgGluGluLysValGlySerGlnSerPheSerGlySer 659
 DB 1943 AGAAGAGTCAAGGGAAGAGTGGGGAAGTGGGAGTCAAGTCTTTCGGCAGCA 2002
 QY 659 e11y11e11e11e11e11e11e11e11e11e11e11e11e11e11e11e11e11e 665
 DB 2003 TGAATATCATGAGGCTTCC 2022
 RESULT 20
 AX921915 2200 bp DNA linear PART 18-DEC-2003
 LOCUS AX921915
 DEFINITION Sequence 255 from Patent WO02068649.
 ACCESSION AX921915
 VERSION AX921915.1 GI:40215409
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1
 AUTHORS
 JOURNAL Patent: WO 02068649-A 255 06-SRP-2002;
 Cuirogen Corporation (US)
 FEATURES
 source
 1..2200
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 ORIGIN
 Alignment Scores:
 Pred. No.: 0 Length: 2200
 Score: 408.00 Matches: 435
 Percent Similarity: 99.328 Conservative: 0
 Best Local Similarity: 99.328 Mismatches: 0
 Query Match: 61.358 Indels: 3
 DB: 6 Gaps: 0
 US-10-029-345A-109 (1-665) x AX921915 (1-2200)
 QY 231 GluLysAlaLysAlaSerAsnGlyCysValLeuValH1sCysLeuAlaGly11eSerArg 250
 DB 793 GAGAAAGCAAAAGCCCTCAATGATGTCTTAAAGTCACTGTTTAACTGGATCTCCCG 852
 QY 251 SerAlaThr11eAla11eAlaTyr11eMetLysArgMetAspMetSerLeuAspGluAla 270
 DB 853 TCGGACCATGCTATCCGCTTACATCATGAAGAGATGAGATGATGATGATGATGATGATGATGAT 912
 QY 271 Tyr--ArgPheValLysGluLysArgProThr11eSerProAsnPheAsnPheLeuGly 289
 DB 913 TACAGAGATTTGTGAAGAAAAGACCTTATATCTCAAACTTCAATTTTCTGGGC 972
 QY 290 GlnLeuLeuAspTyrGluLysLysValLysValAsnGlnThrGlyAlaSerGlyProLysSer 309
 DB 973 CAATCTCTGAGATTAAGAAAGAAATTAAGAAACCAAGACTGAGCATCAAGGCCCAAGAGAC 1032
 QY 310 LysLeuLysLeuLeuH1sLeuGluLysProAsnGluProAlaValSerGluGly 329

TITLE Direct Submission
JOURNAL Submitted (07-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 6 (bases 1 to 172206)
AUTHORS Worley, K.C.
TITLE Direct Submission
JOURNAL Submitted (28-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 7 (bases 1 to 172206)
AUTHORS Worley, K.C.
TITLE Direct Submission
JOURNAL Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT On Nov 19, 2002 this sequence version replaced gi:7658297.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL:
<http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

FEATURES

source

Location/Qualifiers

1..172206

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="12"

/clone="RP11-253119"

176..413

/standard_name="G41293"

repeat_region

complement(888..1190)

/rpt_family="AluSx"

repeat_region

complement(1675..1981)

/rpt_family="AluYb8"

repeat_region

complement(1985..2232)

/rpt_family="AluSx"

repeat_region

complement(2398..2437)

/rpt_family="AluSx"

3010..3016

/function="low quality"

repeat_region 3300..4444 /rpt_family="L1MA8"

repeat_region 4449..4749 /rpt_family="AluUo"

repeat_region 4880..5143 /rpt_family="L2"

repeat_region 5192..5266 /rpt_family="L2"

repeat_region complement(6255..6342) /rpt_family="L1PA7"

repeat_region 7602..7623 /rpt_family="AT_rich"

repeat_region 8096..8126 /rpt_family="(TG)n"

repeat_region complement(8135..8267) /rpt_family="AluUo"

repeat_region 8296..8586 /rpt_family="AluSp"

repeat_region 8911..8938 /rpt_family="AT_rich"

repeat_region complement(10581..10724) /rpt_family="MIR"

repeat_region 10995..11290 /rpt_family="AluUo"

repeat_region 14543..14574 /rpt_family="AT_rich"

repeat_region complement(15322..15625) /rpt_family="AluSx"

repeat_region 17554..17716 /rpt_family="FRAM"

repeat_region 17785..18079 /rpt_family="AluSx"

repeat_region 18231..18527 /rpt_family="AluSp"

repeat_region 18706..18741 /rpt_family="AT_rich"

repeat_region 18901..19190 /rpt_family="AluUo"

repeat_region 19379..19754 /rpt_family="MIR"

repeat_region complement(20289..20337) /rpt_family="MIR"

repeat_region 21165..21658 /rpt_family="L1MA9"

repeat_region complement(22241..22463) /rpt_family="L2"

repeat_region 22638..23279 /rpt_family="L2"

repeat_region complement(24110..24233) /rpt_family="L2"

STS 24233..24361 /standard_name="G62270"

STS 24267..24404

Alignment Scores:

Pred. No.: 0

Score: 394.00

Percent Similarity: 100.00%

Best Local Similarity: 100.00%

Query Match: 59.25%

DB: 9

Length: 172206

Matches: 394

Conservative: 0

Mismatches: 0

Indels: 0

Gaps: 0

US-10-029-345A-109 (1-665) x AC007619 (1-172206)

QY 272 ATGPhEvAllyGlyLylGlyArGProThrilleserProaenPhEaenPhelenglyGlnleu 291

DB 55841 AGATTTCGAAGAAAGAAAAGACCTATATCTCCAAACTTCAATTTCTGGGCCAACTC 55900

QY 222 LeuAaPlyrGlyLylGlyelleylaaanglnthrglyalaserGlyProlysaSerlyeueu 311

DB 55901 CTGAGCTATGAGAAAGATTAAAGAACGATCGAGCATCGGGCCAAAGCAAACTC 55960

QY 312 LyvleuLeuHleuGlyLylGlyrProaenglyProvalProAlaValaserGlyLylGln 331


```

Db      55961 AAGCTCTGCACTGGAGAGCAATGAACTGTCCTCGCTGTCAGAGGGTGGACAG 56020
Qy      332  TysSerGlnThrProLeuSerProProCysAlaAspSerAlaThrSerGlnAlaAlaGly 351
Db      56021 AAAAGGAGACGGCCCTCAGTGCACCTCTGCGACTCTGCTCACTCAAGAGCAGACGAG 56080
Qy      352  GlnArgProValHisProAlaSerValProSerValProSerValGlnProSerLeuLeu 371
Db      56081 CAAAGGCCCTGTCATCCCGCCAGCGTCCCGACGGTCCCGACGGTGCAGCCGTCGCTGTA 56140
Qy      372  GluAspSerProLeuValGlnAlaLeuSerGlyLeuHisLeuSerAlaAspArgLeuGln 391
Db      56141 GAGGACAGCCCGCTGGTACAGCGGCTCAGTGGGCTCCTGTCGTCGACAGAGCGTGGAA 56200
Qy      392  AspSerAsnLeuLeuValArgSerPheSerLeuAspIleLysSerValSerTyrSerAla 411
Db      56201 GACAGCAATAACTCAAGGGTCCCTCTCTGATATCAAAATCAGTTTCATATTCAGCC 56260
Qy      412  SerMetAlaAlaSerLeuHisGlyPheSerSerSerGluAspAlaLeuGlnTyrTyrLys 431
Db      56261 AGCATGGACGATCCTTACATGGCTTCTCTCATCAGAAAGTCTTTGAAATACATCAAA 56320
Qy      432  ProSerThrThrLeuAspGlyThrAsnLysLeuGlnPheSerProValGlnGlnLeu 451
Db      56321 CCTTCACACTCTGGATGGAGCAACAGATATGCGAGTTCTCCCGCTGTCAGGAACTA 56380
Qy      452  SerGlnGlnThrProGlnThrSerProAspArgGlnGlnAlaSerIleProLysLeu 471
Db      56381 TCGAGACACACTCCCAAAACCAAGCTCTGTAAGAGGAAGCAAGATCCCAAGAGCTG 56440
Qy      472  GlnThrAlaArgProSerAspSerGlnSerLysArgLeuHisSerValArgThrSerSer 491
Db      56441 CAGACCGCAGGCGCTTCAGACAGCCAGACCAAGCATTCGCTGCAAGACGACGACC 56500
Qy      492  SerGlyThrAlaGlnArgSerLeuLeuSerProLeuHisArgSerGlySerValGlnAsp 511
Db      56501 AGTGGACACCGCCAGAGGTCCTTTATCTTCACATGATCAGAAAGTGGAGCGTGGAGGAC 56560
Qy      512  AsnTyrHisThrSerPheLeuPheGlyLeuSerThrSerGlnGlnHisLeuThrLysSer 531
Db      56561 AATTACCAACACAGCTCTCTTTTCGCGCTTTCCACACGACACACACCTCCAGAGCT 56620
Qy      532  AlaGlyLeuGlyLeuLysGlyTyrPheHisAspIleLeuAlaProGlnThrSerThrPro 551
Db      56621 GCTGGCGCTGGGCTTAAGGGCTGGCACTCGGATCTTGCCCCCAGACCTCTACCCCT 56680
Qy      552  SerLeuThrSerSerTyrPheAlaThrGlnSerSerHisPheTyrSerAlaSerAla 571
Db      56681 TCCCTGACACAGAGCTGATATTTTGCACAGAGTCTCTCACTTCACTGCTCCAGGC 56740
Qy      572  IleTyrGlnGlySerAlaSerTyrSerAlaTyrSerCysSerGlnLeuProThrCysGly 591
Db      56741 ATCTACGGAGGAGGAGGAGGAGTTCCTGCTACACTGCAGCAGCACTGCCCACTTGGCGGA 56800
Qy      592  AspGlnValTyrSerValArgArgGlnLysProSerAspArgAlaAspSerArgArg 611
Db      56801 GACCAAGTATATCTCTGCGGAGCGCGGAGGAGCCAAAGTGAAGAGCTGACCTCGCGCGG 56860
Qy      612  SerTyrPheGlnGlnSerProPheGlnLysGlnPheLysArgArgSerCysGlnMetGln 631
Db      56861 AGCTGGCATGAAAGAGAGCCCTTTGAAAAAGCAGTTTAAACGAGAAAGTGCACAAATGGA 56920
Qy      632  PheGlyGlnSerIleMetSerGlnAsnArgSerArgGlnGlnLysValValGlySer 651
Db      56921 TTTGAGAGAGCATATGTCAAGAAAGGTCACGGGAAAGACTCTGGGAAAGTGGGCGAGT 56980
Qy      652  GlnSerSerPheSerGlySerMetGlnIleIleGlnValSer 665
Db      56981 CAGTCTAGCTTTTTCGGGCGAGCATGGAATCATTTAGTCTCC 57022

```

RESULT 22
AC131617/c

```

LOCUS      AC131617      188344 bp      DNA      linear      HTG 17-SEP-2002
DEFINITION Homo sapiens clone RP11-613M22. *** SEQUENCING IN PROGRESS ***, 7
ACCESSION  AC131617
VERSION    AC131617.3 GI:22953886
KEYWORDS   HTG; HTGS PHASE1; HTGS ACTIVEFIN.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 188344)
            Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-ouman,F.R., Allen,C.,
            Albrooks,S.L., Amarantunga,H.C., Are,J.R., Ayale,M., Banks,T.,
            Barbarta,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Bivrea,M., Brown,E., Brown,M., Bryant,N.P.,
            Bunay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Kocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harrie,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Huliyil,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Krstovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louised,H.,
            Lorado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheswari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E.,
            Maesey,E., Mawhinney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokweto,S., Ogunu,M., Okunonu,G.,
            Orsgunye,N., Oyadeo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primm,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rivers,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G.,
            Scherer,S., Scott,G., Shen,H., Shoochitari,N., Sisson,I.,
            Sodergren,E., Sotaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Swatek,A., Taborp,P., Tamerisa,A., Tameiras,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,
            Umanai,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
            Direct Submission
            Unpublished
            2 (bases 1 to 188344)
            Worley,K.C.
            Direct Submission
            Submitted (25-AUG-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
            3 (bases 1 to 188344)
            Worley,K.C.
            Direct Submission
            Submitted (17-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
            On Sep 17, 2002 this sequence version replaced gi:22538362.
            ----- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Drafting Center Code: BCM
            Contact: hgsc-help@bcm.tmc.edu
            ----- Project Information

```


REMARK

Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, NIH-MGC Project URL: <http://mgc.nci.nih.gov>
 Contact: MGC help desk
 Email: cgabbs-remail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: amg@bcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M., Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
 Series: IRAX Plate: 51 Row: n Column: 21
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

FEATURES

source

CDS

Location/Qualifiers
 1..2807
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:5176724"
 /issue_type="Brain, Lung, Testis, adult, pooled whole"
 /clone_lib="NIH MGC_115"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 <1..1852
 /codon_start=2
 /product="Unknown (protein for IMAGE:5176724)"
 /protein_id="AAH31643.1"
 /db_xref="GI:21594973"
 /translation="CSKMKRRLOQDKVITLTHSAKHKIDDCSKVNVYDQSSQDVASLSDCFVLVLGKLEKSPNSVHLAAGFAFSRCPGLCEBKSFLVPCISQCLPVAINGPRLIPNLVYLGCDVNLKELMQNGIGVLAASATCEKPFIPSPHFLRPVNDSECEKILPMLDKSVDFIEKAKASNCVLAIGASRATIAIAYIMKRMMSLDEAYRFEKKEKRLTSPNPNVLGOLDYEKKIKNOTGASGPKSLKLHLKRNEMVPAVSEGGKSETPLSPCADSAETSEAGORPVHAPSVSPVSPSLSEDSPLQALSGIHSAARLEBSNKLKRSFSLDITKSVSYASMAASLHGFSSSEDLATYRSTLIDSTNKLQFSPVDELSEQTPEISPDKEKASI PKLQTPARPSDQSKLHVRSNQTARSL LSPHRSQGVEDNHTSPLFGISTQOHLTKSAGLKGMSHDILAPOTSTPLTSWSVYATESSHRYSAIAIYGGASYSAYSCSLPTCGDQVYSVRROKPSRADRSRMHZEESPFEKQFRKRCQMEFGESIMSENRSRRELKGVSGSFSGSMETIEVS"

ORIGIN

Alignment Scores:

Pred. No.: 0 Length: 2807
 Score: 371.00 Matches: 613
 Percent Similarity: 99.03% Conservative: 0
 Best Local Similarity: 99.03% Mismatches: 3
 Query Match: 55.79% Indels: 6
 DB: 9 Gaps: 0

US-10-029-345A-109 (1-665) x BC031643 (1-2807)

QY 50 CysSerIysIeuMerIyAhrGargLeuGInGInaApIySValIeuIleThrGluIeuIle 69
 DB 2 TcCTCAAGCTTATGAAGGAAAGGTTGCAACAGCAAGAGTGTATTATTAACAAGCTCAATC 61
 QY 70 GInHISerAlaIyehIIsySValAspIleAspCysSerGInIySValIyValIyTyrAsp 89
 DB 62 CAGCAATTACGCAAAACATTAAGTTGACATTGATTCGACAGTCAGAGGTTGATTACAT 121
 QY 90 GInSerSerGInaSpValAlaSerIeuSerSerAspCysPheLeuThrValIeuIeuGly 109
 DB 122 CAAAGCTCCCAAGATGTTGCTCTCTCTTCAAGCTGTTTCTCACTGTAATCTTGAGT 181

QY 110 LysIeuGluIySserPheAsnSerValIhiSLeuLeuAlaGlyIyPheAlaGluPheSer 129
 DB 182 AAACCTGGAGAAAGCTTCAACCTGTTCACTGCTGGAGGGGTTGGTAGTCTCT 241
 QY 130 ArGcYsPheProGlyIySLeuGluGluIyLysSerThrIeuValProThrCysIleSerGln 149
 DB 242 CATTGTTTCCCTGGGCTCTGTGAGAGAAATCCACTAGTCTTCACTGCAATTTCTGAG 301
 QY 150 ProCysIeuProValAlaAsnIleGlyProThraGlyIleuProAsnIeuTyrIeuGly 169
 DB 302 CCTTCACTCACTGTTGGCAACATTGGGCCAACCCGAATTTCTTCCCAATCTTATCTTGCC 361
 QY 170 CysGInaGAspValIeuAsnIySLeuIleGlnIySValIyValIyValIyValIyValIy 189
 DB 362 TGGCAGCAGAGATGTCCTCAACAGAGAGCTGAT-GCAGAGATGGAGATGGTATAGTGT 420
 QY 189 uAsnAlaSerTyr-ThrCysProIySProAspPheIleProGluSerHisPheLeuArgy 209
 DB 421 AAATCCAGCAA-TACCTGTCCAAAGCTGACTTATCTCCGAGTCTCATTTCTTGGCTG 479
 QY 209 aIProValAsnAspSerPheCysGluIySileuProTyrIeuAspIySserValAsp 229
 DB 480 TGCCTGTAATGACAGCTTTTGTGAGAAATTTTCCGTGGTGGACAAATCAGTAGATT 539
 QY 229 heIleGluIyAlaIyAlaSerAnGlyCySValIeuValIhiSyleuAlaGlyIyIleS 249
 DB 540 TCATTGAGAAAGCAAAAGCTTCCCAATGATGATGTTAGTCACTGTTTACCTGGAGTCT 599
 QY 249 eArGSeAlaThrIleAlaIleAlaTyrIleMetIyArGMetAspMetSerIeuAspG 269
 DB 600 CCCGCTCCGCCACCATGCTTATCCGCTACATCAAGAGAGATGACATGCTTTAGATG 659
 QY 269 lualATyArGpheValIySgluIySArgProThrIleSerProAsnIeuAsnIeuIeuG 289
 DB 660 AAGCTTACAGATTGTGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 719
 QY 289 lYgInIeuIeuAspTyrGluIySylIySValIySValIySValIySValIySValIySVal 309
 DB 720 GCCAATCTCTGACTGATGAGAAAGAAAGATTAAGAAACAGACCTGGAGCAACAGGCAAG 779
 QY 309 eTlYsLeuIyLeuIeuIhiSLeuGluIySProAnGluProValProAlaValSerGluG 329
 DB 780 GCMAATCAACCTGCTGACCTGAGAGAGCAAAATGAACCTGCTGCTGCTGCTGCTGAGG 839
 QY 329 lYgIyGInIySserGluThrProIeuSerProProCysAlaAspSerAlaThrSerGluA 349
 DB 840 GTGACAGAGAAAGGAG 899
 QY 349 lAaIaGlyGlnArgProValIhiSProAlaSerValProSerValGlnProS 369
 DB 900 CAGCAGAGCAAAAGGCGGTGATCCCGCAGCGAGGCGGCGGCGGCGGCGGCGGCGGCGG 959
 QY 369 eTlLeuLeuGluAspSerProIeuValGlnAlaIeuSerGlyIyLeuIhiSLeuSerAlaAsp 389
 DB 960 CGCTGTTAG 1019
 QY 389 rGLeuGluAspSerAsnIySLeuIySArgSerPheSerIeuAspIleIySValIserT 409
 DB 1020 GGCTGGAAGACAGCAATTAAGCTCAAGGCTTCTCTCTGATATCAAAATCAATTCAT 1079
 QY 409 YrSerAlaSerMetAlaIaSerIeuIhiSglyPheSerSerSerGluAspAlaIeuGluT 429
 DB 1080 ATTGAGCAGAGATGAGAGCACTTCAATAGCTTCTCTCAATCAAGAGATGCTTTGGAAT 1139
 QY 429 YrTyrIySProSerThrThrIeuAspGlyTyrAsnIySLeuCySglIyPheSerProValG 449
 DB 1140 ACTACAAACCTTCACTACTGATGGAGACCAACAGTATGCAATTCCTCCCTGTTCC 1199
 QY 449 InGluIeuSerGluGlnThrProGluThrSerProAspIySgluIyAlaSerIleProL 469
 DB 1200 AGGAATATATGAGAGACAGACTCCGAAACCAAGTCTGTATTAAGAGAGAGAGAGAGAG 1259

QY 469 yslvleuglnThrAlaArgProSerAspSerGlnSerIyValGluHisSerValArgT 489
 Db 1260 AGAGCTGAGACCGCAGGCTTTCAGACGACGACGACGATTCGATCGGTAGAA 1319
 QY 489 hSerser-SerGlyThrAlaGlnArgSerLeuSerProLeuHisArgSerGlySer 508
 Db 1320 CCGACAA-CAGTGGCACCCGCCGAGAGTCCCTTTATCTCCATCGATCGAATGGGAGC 1378
 QY 509 ValGluAspAsnThrIleThrSerPheLeuPheGlyLeuSerThrSerGlnGlnHisLeu 528
 Db 1379 GTGGAGACAAATTAACAACACGACCTTCTTGGCCCTTCCACGACGACGACCTTC 1438
 QY 529 ThrIySerAlaGlyLeuGlyLeuGlySerGlyTrpHisSerAspIleLeuAlaProGlnThr 548
 Db 1439 AGCAATCTGCTGGCTGGCCCTTAAGGCGTGGCATCTCGATCTTGGCCCCCAGACC 1498
 QY 549 SerThrProSerLeuThrSerSerTrpIyPheAlaThrGluSerSerHisPheIySer 568
 Db 1499 TCTACCCCTCCCTGACGACGACGACCTGTATTTTGCACAGAGCTTCACACTTCTACTCT 1558
 QY 569 AlaSerAlaIleThrGlyGlySerAlaSerIySerAlaIySerCySerGlnLeuPro 588
 Db 1559 GCTTCAGCATCTACGAGGACGACGACGACGACGACGACGACGACGACGACGACGAC 1618
 QY 589 ThrCyGlyAspGlnValIySerValIySerValIySerValIySerValIySerValIySer 608
 Db 1619 ACTGGCGAGACCAAGTCTTCTTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1678
 QY 609 SerArgArgSerTrpHisGluGluSerProPheGluLeuGlnPheIyValArgArgSerCy 628
 Db 1679 TCGCGCGGAGCTGCGATGATGAAGAGAGCCCTTGAAGAAGAGAGAGAGAGAGAGAGAG 1738
 QY 629 GlnMetGluPheGlyGlyLeuSerIleMetSerGluAsnArgSerArgGluGluLeuGly 648
 Db 1739 CAATGAATTTGGAGAGACATCATGTCAAGAAAGAGACGACGAGAGAGAGAGAGAGAG 1798
 QY 649 ValGlySerGlnSerSerPheSerGlySerMetGluIleIleGluValSer 665
 Db 1799 GTGGGAGTCACTGCTTTCGCGGACGATGGAATCATTTAGGTTCC 1849
 RESULT 24
 AX482372 5111 bp DNA linear PAT 16-AUG-2002
 LOCUS AX482372 Sequence 41 from Patent WO02057460.
 DEFINITION AX482372
 ACCESSION AX482372
 VERSION AX482372.1 GI:22316945
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE
 1 Todderud,C.G., Bol,D., Finger,J., Lee,L., Nelson,T., Schieven,G.,
 Suchard,S., Banas,D., Bassolino,D., Feder,J., Kyslek,S.,
 Mcatee,P., Mintier,G., Siemers,N., Jackson,D.G. and Ramanathan,C.
 TITLE Polynucleotides encoding human phosphatases
 JOURNAL Patent: WO 02057460-A 41 25-JUL-2002;
 SQUIBB BRISTOL MYERS CO (US)
 FEATURES
 source location/Qualifiers
 1. 5111
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 ORIGIN
 Alignment Scores:
 Pred. No.: 0 Length: 5111
 Score: 350.00 Matches: 661
 Percent Similarity: 98.80% Conservative: 0
 Best Local Similarity: 98.80% Mismatches: 4
 Query Match: 52.63% Indels: 8
 DB: 6 Gaps: 0

US-10-029-345A-109 (1-665) x AX482372 (1-5111)
 QY 1 MetAlaHisGluMetIleGlyThrGlnIleValThrGluArgValAlaLeuLeuGlu 20
 Db 470 ATGGCCCATGAATGATGTTGAATCTCAATTTGATGAAAGGTGGTGGCTGCTGGAA 529
 QY 21 SerGlyThrGluValLeuLeuIleLeuSerArgProPheValGluIyTrpAsnThrSer 40
 Db 530 AGTGAACCGAAGAAAGTGTGCTAATGATATACCGGCCCATTTGTGAAATACATATC 589
 QY 41 HisIleLeuGlnAlaIleAsnIleAsnCySerIyLeuMetIyValArgArgLeuGlnGln 60
 Db 590 CCAATTTGGAAGCCATTAATATCAACTCTCCAGCTTATGAAAGGAGGTGGCAACAG 649
 QY 61 AspIyValLeuIleThrGluLeuIleGlnHisSerAlaIyHisIyValAspIleAsp 80
 Db 650 GACAAAGTGTAAATTAACAAGCTCATCCGACATTCAGCAAAATTAAGGTGACATTTAT 709
 QY 81 CySerGlnIyValValIyValIyTrpAspGlnSerSerGlnAspValAlaSerLeuSer 100
 Db 710 TGCAGTCAGAGAGTGTGATTCATGATCAAAAGCTCCCAAGATGTTGCTCTCTCTTCA 769
 QY 101 AspCyPheLeuThrValLeuLeuGlyIyValLeuGlyIyValSerPheAsnSerValHisLeu 120
 Db 770 GACTGTTTCTCACTGTACTTCTGGGTAAACTGAGAAAGAGCTTCACTGTTCACTG 829
 QY 121 LeuAlaGlyIyPheAlaIleLeuPheSerArgCyPheProGlyLeuCyArgGluGlySer 140
 Db 830 CTTCAGAGGTGGTGGTGTGATGTTCTCTGTTTTCCTGCGGCTCTGTAAAGGAAATCC 889
 QY 141 ThrLeuValProThrCyHisIleSerGlnProCySleuProValAlaAsnIleGlyProThr 160
 Db 890 ACTCTAGTCTCACTGCTGCTTCTGAGCTTCTGCTTCTGCTGCTGCTGCTGCTGCTGCT 949
 QY 161 ArgIleLeuProAsnLeuIyLeuGlyCyGlnArgAspValLeuAsnIyGluLeu 180
 Db 950 CCAATTTCTCCCAATTTTATCTTGGCTCCAGCGAGATGTCTCAACAAGAGCTGAT- 1008
 QY 181 -GlnGlnAsnGlyIleGlyIyValValLeuAsnIleSerIyTrpThrCyProIyPheAsp 200
 Db 1009 GAGACGAATGGATGGATGTTATGTGTTAATGCCAGCAA-TACCTGTCCAAAGCTGTACT 1067
 QY 200 IleIleProGluSerHisPheLeuArgValProValAsnAspSerPheCyGluIyIle 220
 Db 1068 TTAATCCCGAGTCTCATTTCTGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1127
 QY 220 euProTrpLeuAspIySerValIyPheIleGluValAlaIyValAsnAsnGlyCyArg 240
 Db 1128 TCCCGGTGGAGCAATCAGTAGATTTTCAATGAGAAACCAAGCCTCAATGATGATG 1187
 QY 240 AlLeuValHisCySleuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaIyIle 260
 Db 1188 TTTCTGTGACATGTTTATGCTGGAGATCTCCGCTCCGCCACATCGCTATCGCTATCATCA 1247
 QY 260 eIyValArgMetAspMetSerLeuAspGluAlaIyTrpArgPheValIyGluIyArgPro 280
 Db 1248 TGAAGAGATGAGCATGCTTTTATGATGAAGCTTACAGATTTGTAAGAAAGAAAGACCTTA 1307
 QY 280 hIleSerProAsnPheAsnPheLeuGlyGlnLeuLeuAsp-TyGluIyIyIyIyIy 299
 Db 1308 CTATATCTCAAACTTCATTTTCTGGGCCAACTCTGGC-CTATGAGAAAGATTAAG 1366
 QY 300 AsnGlnThrGlyAlaSerGlyProIySerIyValLeuIyLeuLeuHis-LeuGluIyAsp 319
 Db 1367 AACCAAGCTGAGCATCAGGCGCAAGCAAACTCAACTCACTCTGTC-CTGTGAGAGAGCC 1425
 QY 319 AsnGluProValProAlaValSerGluIyGlyGlnIySerGluThrProLeuSerP 339
 Db 1426 AATATGAACCTGCTCTGCTGCTCTCAAGAGGTGACGAAAGGAGAGAGAGAGAGAGAGAG 1485
 QY 339 oProCyValAspSerAlaThrSerGluAlaIyGlnArgProValHisProAlaSe 359
 Db 1486 ACCCTGTGCGACTGCTGCTACCTCAGAGGACAGACAAAGGCGGTGATCCGCGCAG 1545

QY 359 rValProSerValProSerValGlnProSerLeuLeuGluAspSerProLeuValGlnAl 379
 DB 1546 CGGCGCCAGGCTGCCAGCTGCAGCCGCTGCTTTAGAGACAGCCCGCTGTACAGGC 1605
 QY 379 aLeuSerGlyLeuHisLeuSerAlaAspArgLeuGluAspSerAsnLysLeuLysArgSe 399
 DB 1606 GCTCAGTGGGCTGCACCTGTCCGACAGACAGCTGGAAGACAGCAATATAGCTCAAGGCTTC 1665
 QY 399 rPheSerLeuAspLysLeuSerValSerTyrSerAlaSerMetAlaLysLeuHisGln 419
 DB 1666 CTCTCTCTGGATATCAAAATCAATTCATATTCAGCAGAGCTGCAGACATCTTATCATG 1725
 QY 419 yPheSerSerSerGluAspAlaLeuGluLysTyrLysProSerThrThrLeuAspGlyLys 439
 DB 1726 CTCTCTCTCATCGAAGATGCTTTGGAATATCAAAATCTTCCACTACTCTGTGATGGAGC 1785
 QY 439 rAsnLysLeuGlyGlnPheSerProValGlnGluLeuSerGluGlnThrProGluThrSe 459
 DB 1786 CAACAAGCTATGCGCAGTTCTCCCTGTTCCAGAACTATCGAGCAGACTCCGAAACAG 1845
 QY 459 rProAspLysGluGluAlaSerLysLeuProLysLysLeuGlnThrAlaArgProSerAspSe 479
 DB 1846 TCTGTATTAAGAGAAAGCCAGACATCCCAAGAACTGCAGACCGCAGGCTTCAGACAG 1905
 QY 479 rGlnSerLysArgLeuHisSerValArgThrSerSerSerSerGlyThrAlaGlnArgSerLe 499
 DB 1906 CCAGACCAAGCATTCATTCGCTCGAACCAGACAGCTGCGCAGCCGCAAGATCCCT 1965
 QY 499 uLeuSerProLeuHisArgSerGlySerValGluAspAsnTyrHisThrSerPheLeuAsp 519
 DB 1966 TTTATCTCCACCTGCATCGAAGTGGAGCGTGGAGGCAATTAACAACAGCTTCTTT 2025
 QY 519 eGlyLeuSerSerThrSerGlnGlnHisLeuThrLysSerAlaGlyLeuGlyLeuLysGlyTr 539
 DB 2026 CGGCTTTTCCACAGCCAGCAGACCTCAAGAACTGCTGCGCTTGAAGGCGT 2085
 QY 539 pHisSerAspLysLeuAlaProGlnThrSerThrProSerLeuThrSerSerTyrTrp 559
 DB 2086 GCACCTGGATATCTTGCGCCGCCACAGACCTTACCCCTTCCGACACAGAGCTGTATTT 2145
 QY 559 eAlaThrGlnSerSerHisPheTyrSerAlaSerAlaLysTyrGlyLysSerAlaSerTyr 579
 DB 2146 TCCCAAGAGTCTCTACACTTCTGCTGCTGACGACATCAAGAGGCAAGTGCAGTTA 2205
 QY 579 rSerAlaTyrSerCysSerGlnLeuProThrCysGlyAspGlnValTyrSerValArgAr 599
 DB 2206 CTCTGCTTCAAGCTGCAGCAGCTGCACCTTGCAGAGACCAAGCTATTTCTGTGCGAG 2265
 QY 599 gArgGlnLysProSerAspArgAlaAspSerArgArgSerTrpHisGlnGluSerProPh 619
 DB 2266 GCGGCGAAGCCAGATGACAGAGCTGACCTGCGGGGAGCTGGCATTGAAGAGGCCCTT 2325
 QY 619 eGluLysGlnPheLysArgArgSerCysGlnMetGluPheGlyGluSerTyrLysSerGln 639
 DB 2326 TGAAGAAGAGTTTAAACGCAAGAGCTGCCAAATGGAATTTGAGAGACATCATGTGAGA 2385
 QY 639 uAsnArgSerArgGluGluLeuGlyLysValGlySerGlnSerSerPheSerGlySerMe 659
 DB 2386 GAACAGGTACCGGAGAGAGCTGGGAAAGTGGGAGCTGACGTCTTACCTTTTGGGAGAGAT 2445
 QY 659 rGluLeuLeuGluValSer 665
 DB 2446 GGAATTCATTGAGTCTCC 2464
 RESULT 25
 AX921917 2071 bp DNA linear PAT 18-DEC-2003
 LOCUS AX921917
 DEFINITION Sequence 257 from Patent WO02068649.
 ACCESSION AX921917
 VERSION AX921917.1 GI:40215410
 KEYWORDS
 SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1
 AUTHORS Patent: WO 02068649-A 257 06-SEP-2002;
 JOURNAL Curegen Corporation (US)
 FEATURES location/Qualifiers
 source 1..2071
 /organism="Homo sapiens"
 /mol_type="unassigned DNA"
 /db_xref="taxon:9606"
 ORIGIN
 Alignment Scores:
 Pred. No.: 1,94e-288 Length: 2071
 Score: 304.00 Matches: 658
 Percent Similarity: 98.65% Conservative: 0
 Best Local Similarity: 98.65% Mismatches: 7
 Query Match: 45.71% Indels: 9
 DB: Gaps: 0
 US-10-029-345A-109 (1-665) x AX921917 (1-2071)
 QY 1 MetaLHisGluMetLysGlyThrGlnLysValThrGluArgLeuValAlaLeuLeuGlu 20
 DB 61 ATGGCCCATGAGATGATTTGGAATCTCAATTTGTTACTGAGAGTGTGCTCTCTGGA 120
 QY 21 SerGlyThrGluLysValLeuLeuLysAspSerArgProPheValGluTyrAsnThrSer 40
 DB 121 AGTGAACGGAAGAAAGTCTCTTAATTGATGATCCGCGCATTTGGAATCAATATCATCC 180
 QY 41 HisLysLeuGluAlaLysAsnLysAsnLysSerLysLeuMetLysArgLeuGln 60
 DB 181 CACATTTTGAAGCATTAATATATCACTGCTCCAAAGCTTATGAAGCAAGGTTGCAACAG 240
 QY 61 AspLysValLeuLysThrGluLeuLysGlnHisSerAlaLysHisLysValAspLys 80
 DB 241 GACAAAGTTTAATTAACAGCTCATTCAGCATTCACGAAACATAAGTTGACATTGAT 300
 QY 81 CysSerGlnLysValValValTyrAspGlnSerSerGlnAspValAlaSerLeuSerSer 100
 DB 301 TGCAGTCAGAAAGTGTGATTTAGATCAAGCAAGCTCCCAAGATGTGCTCTCTTCA 360
 QY 101 AspPheLeuThrValLeuLeuGlyLysLeuGlnLysSerPheAsnSerValHisLeu 120
 DB 361 GACTGTTTCTCACTGATCTTCTGGTAAACTGGAAGAGCTTCAACTGTTCACCTG 420
 QY 121 LeuAlaGlyGlyPheAlaGluPheSerArgCysPheProGlyLeuCysGluGlyLysSer 140
 DB 421 CTTCAGAGTGGGTTGCTGAGTTCTCTGCTTTTCCCTGCTCTGTGAAGAAATCC 480
 QY 141 ThrLeuValProThrCysLysSerGlnProCysLeuProValAlaAsnLysGlyProThr 160
 DB 481 ACTCTAGTCCCTCACTGCAATTCCTCAGCTTGTGCTTCACTGTGCAACATTTGGGCCAAC 540
 QY 161 ArgLysLeuProAsnLeuTyrLeuGlyCysGlnArgAspValLeuLysGlyLeuLys 180
 DB 541 CGAATTTCTCCCAATCTTATCTTGTGCTGCACAGAGATGCTTCAACAAGAGCTGAT- 599
 QY 181 -GlnGlnAsnGlyLysGlyTyrValLeuAsnAlaSerTyr-ThrCysProLysProAsp 200
 DB 600 GCAGCAAGATGGGATGTGATATGTATATGCAAGAA-TACTGTCCAAACCTGACT 658
 QY 200 HeLeuProGluSerHisPheLeuArgValProValAsnAspSerPheCysGlyLysLys 220
 DB 659 TTAATCCCGAGTCTCAATTTCTCTGCTGCTGCTGATGACACACTTTTGAAGAAATTT 718
 QY 220 euProThrLeuAspLysSerValAspPheLysGlnLysAlaLysArgGlyCysArg 240
 DB 719 TGCCTGTGTGCAAAATCAAGATTTCAATTAGAAAGCAAGGCTCCAAATGAGATGTG 778
 QY 240 alleuValHisCysLeuAlaGlyLysSerArgSerAlaThrLysAlaLysTyrLys 260

```

Db      779 TTCTATGACATCTTTAGCTGGGATCTCCCGCTCCGCCACATCGCTATCGCTTACATCA 838
Qy      260 eelyaarghweapmetserleuapglua1a7yarghveallysglulyaargprot 280
Db      839 TGAAGAGGATGACATGCTTTAGATGAAAGCTTACAGATTGTGAAAGAAAAGACCTTA 898
Qy      280 hrliserproanpheanpheleuglyglntleuaperyrglyulyalya11alya 300
Db      899 CTATATCTCCAACTTCAATTTCTGCGCCACTCTGACCTATGAGAAAGAAATTAAGA 958
Qy      300 snlgnthrghyalaaserghyprolyserleuyleuyleuyleu11alya11alya 320
Db      959 ACCAGACTGAGATCATGGGCCAAAGACAACTCAAGCTGCTGACCTGGAAAGCCAA 1018
Qy      320 snlgnthrghyalaaserghyprolyserleuyleuyleuyleu11alya11alya 340
Db      1019 ATGAACCTGCTCCTGCTGCTCAGAGGCTGAGAGAAAGAGAGAGAGAGAGAGAGAG 1078
Qy      340 rocyala1aapsera1a1thrserghyala1a1a1a1a1a1a1a1a1a1a1a1a 360
Db      1079 CCGTGTCCGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1132
Qy      360 a1proservalproserval1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a 380
Db      1133 ---CC-AGCTTACCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1189
Qy      380 euserghy1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a 400
Db      1190 TCAGTGGGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1249
Qy      400 heserleuapleuyleuyleuyleuyleuyleuyleuyleuyleuyleuyleuyle 420
Db      1250 TCTCTGTGATATCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 1309
Qy      420 heserleuyleuyleuyleuyleuyleuyleuyleuyleuyleuyleuyleuyle 440
Db      1310 TCTCTCATCAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1369
Qy      440 snlyleuycyglntleuyleuyleuyleuyleuyleuyleuyleuyleuyleuyle 460
Db      1370 ACAAGCTATCCGATCTCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1429
Qy      460 roaaplysglulua1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a 480
Db      1430 CTGATTAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1489
Qy      480 lnsertysargleu1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a 500
Db      1490 AGAGCAAGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1549
Qy      500 euserproleu1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a 520
Db      1550 TATCTCCACGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCG 1609
Qy      520 lyleu1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a 540
Db      1610 GCGTTTCCACACGATCGATCGATCGATCGATCGATCGATCGATCGATCGATCG 1669
Qy      540 lnsertysargleu1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a 560
Db      1670 ACTCGGATATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1729
Qy      560 lathrgluserserghy1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a 580
Db      1730 CCAAGAGCTCTCACTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1789
Qy      580 era1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a 600
Db      1790 CTCCTCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1849
Qy      600 rglntlyserprosera1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a 620

```

```

Db      1850 GCGAGAACCAAGTACAGAGCTGACTCGCGCGAGAGCTGGCATGAAGAGAGCCCTT 1909
Qy      620 lulye1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a 640
Db      1910 AAAAGCATTTTAAACCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1969
Qy      640 snlgnthrghyalaaserghyprolyserleuyleuyleuyleuyleuyleuyle 660
Db      1970 ACGGTCAAGGAGAGAGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2029
Qy      660 lulye1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a 680
Db      2030 AAATCATGAGGTCTCC 2046

RESULT 26
BC042101
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
1 (bases 1 to 3284)
Strausberg,R.D., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,E.J., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buettow,K.H., Scheffer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Heich,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Scheetz,T.E., Brownstein,M.J., Uebli,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullany,S.J., Bosak,S.A., McEwan,P.J.,
Mckernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Morley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Kettelman,M., Madan,A., Rodrigues,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakeley,R.W., Touchman,D.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalsky,U., Smallu,D.E.,
Schnerch,A., Schein,U.E., Jones,S.J., and Marra,M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
JOURNAL
MEDLINE
PUBMED
2 (bases 1 to 3284)
Strausberg,R.
Direct Submission
Submitted (23-DEC-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550,
USA
REMARK
COMMENT
NIH-MGC Project URL: http://mgc.ncl.nih.gov
Contact: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
Web site: http://www-ehg.stanford.edu
Contact: (Dickson, Mark) mdcpaxi1.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
Series: IRAX Plate: 88 Row: a Column: 17

```


QY 440 anlyseuCyvsglnpbeserProvalglnluteusergluglnthrpProgluThrserp 460
DB 1641 ACAAGCTATCCAGTCTCCCTGTTCCAGAACTATCCAGAGCATCCCGAAACCAAGTC 1700
QY 460 roaaplyvsglualaaserileprolyvlyseuglnthralaargProserAspserg 480
DB 1701 CTGATAAAGGAGGAGCAGCATCCCAAGAGCTGCAGACCGGCGCTTTCAGACAGCC 1760
QY 480 InseiyasargleuHiaserValArgThrsersererglyThraaglInaArgserleu 500
DB 1761 AGGCAAGCAGATTGTCATTCGTGCAGAACCCAGAGCTGGCACCAGCCCAAGGTCTCCCTT 1820
QY 500 euserProleuHiaserSerglySerValgluAspAntyrHiaserPheleuPheg 520
DB 1821 TATCTCCACTGCTCATCGAAGTGGAGCCTGAGAGCAATTCACACACAGCTTCTTTTCG 1880
QY 520 lyeuserThrserglnInHiasleuThrsySerValaglYleuglyleuYlYtrpH 540
DB 1881 GCCTTCCACACAGCAGACACCTCAAGAGTCTGGCTGGGCTTAAGGGCTGGC 1940
QY 540 liserAspilleuAlaProglInThrserThrsProserleuthrserserTyrPhea 560
DB 1941 ACTCGATATCTGGCCCCCAGACCTCTACCCCTTCCCTGACAGCAGCTGGATTTTG 2000
QY 560 lathrgluserSerHiaspHiaserValaierYrglyYglYserAlaSerTyrS 580
DB 2001 CCAAGAGTCTCACTTACTCTGCTCAGCACTAGAGAGGAGGAGGAGGAGGAGTACT 2060
QY 580 ezalYrSerCyvserGlnleuProThrCyvsglyAspGlnValYrSerValaArgAsp 600
DB 2061 CTGCTTACAGCTCAGCAGCAGCTGCCACTTGGGAGACCAAGCTATTCTGTCGCGAGC 2120
QY 600 rglInlyvProserAspArgAlaAspSerArgHiaserTyrHiasgluInserProPheg 620
DB 2121 GGCAGAGAGCCAGATGACAGAGCTGACTCGCGCGAGCTGGCATGAAGAGCCCTTTG 2180
QY 620 lulyseglInPheylvAspArgSerCyvsglnMetgluPheglYgluserIleMetSerGln 640
DB 2181 AAAAGCATTTAAAGCAGAAAGCTGCCAAATGAAATTTGAGAGAGATCATATGACAGA 2240
QY 640 snArgserArggluglnleuGlylYvAlglYserGlnserSerPheSerSerglySerMetC 660
DB 2241 ACAGGTACAGGAGGAGAGCTGGGAAAGTGGGAGTACGTACTCTTTTGGGAGAGCATG 2300
QY 660 lulellegluValSer 665
DB 2301 AAATCATTTGAGTCTCC 2317

RESULT 27
AX835305 1916 bp DNA linear PAT 15-DEC-2003
LOCUS Sequence 2429 from Patent EP1347046.
ACCESSION AX835305
VERSION AX835305.1 GI:39921440
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS Iseogi,T., Sugiyama,T., Otsuki,T., Wakamatsu,A., Sato,H., Ishii,S.,
Yamanoto,J.T., Isono,Y., Hio,Y., Otsuka,K., Nagai,K., Irie,R.,
Tamechika,I., Seki,N., Yoshikawa,T., Otsuka,M., Nagahara,K. and
Masuno,Y.
TITLE Full-length cDNA sequences
JOURNAL Patent: EP 1347046-A 2429 24-SEP-2003;
Research Association for Biotechnology (JP)
FEATURES
Source
1..1916
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"

ORIGIN

Alignment Scores:
Pred. No.: 1,42e-164 Length: 1916
Score: 178.00 Matches: 278
Percent Similarity: 99.29% Conservative: 0
Best Local Similarity: 99.29% Mismatches: 1
Query Match: 26.77% Indels: 2
DB: 6 Gaps: 0

US-10-029-345A-109 (1-665) x AX835305 (1-1916)

QY 387 AlAspArgLeuGluuAspSerAnlyseuLyAspSerPheSerleuAspIleYsSer 406
DB 142 GCAGAGAGCTGGAGACAGCAATTAAGCTCAAGCGTCTCTCTCTGATATCAATCA 201
QY 407 ValSerTyrSerAlaSerMetAlaAspSerleuHiasglYpHeserserSergluAspAla 426
DB 202 GTTTCATATTACAGCCAGCATGGCAGCATCTTAACAATGAGCTTCTCATCAGAAAGATGCT 261
QY 427 leuGluTyrTyrlyvProserThrThraAspGlyThraAspLyseuCyvsglnPheSer 446
DB 262 TTGAAATTAACAACCTTCACTTCACTGATGGAGCAACCAAGCTATGCGAGTCTCC 321
QY 447 ProvalglnluteusergluglnthrpProgluThrsProasplysgluAlaSer 466
DB 322 CTTGTTCAAGAACTATCGAGAGAGCTCCGAAACCAAGTCTGATAGAGAGAGAGCAGC 381
QY 467 lIeProlyvlyleuGlnthralaargProserAspSerglnserlyAspArgleuHiaser 486
DB 382 ATCCCAAGAGGTGAGAGCCGCAAGCTTTCAGACAGCAGCAAGAGGATGATTCGATTCG 441
QY 487 ValArgThrserserSerglyThraaglInaArgserleuLeuSerProleuHiasArgSer 506
DB 442 GTCCAGAACAGAGCAGAGTGGCAACCGCCAGAGTCCCTTTTATCTCACTGATCGAAGT 501
QY 507 GlySerValgluAspAntyrHiaserPheleuPheglYleuserThrsSerglnGl 526
DB 502 GGGAGGTGAGAGCAATTAACAACAGCTTCTTTGGCTTTCAT-CAGCCAGCA 560
QY 526 nhIseuThrsySerAlaaglYleuGlyleuYsglyTyrHiaserAspIleleuAlaP 546
DB 561 GCACCTCAGAAAGTGTGCTGGCTGGCTTAAGGGCTGGCACTCGATATCTTGGCCCC 620
QY 546 oglnThrserThrsProserleuthrserserTyrPheAlaThrgluserSerHiasP 566
DB 621 CCAGACTTACCCCTTCCCTGACCAAGCAGCTGGATTTTGGCAGAGATCTTCACACTT 680
QY 566 eTyrSerAlaSerAlaIleTyrGlyYserAlaSerTyrSerAlaTyrSerCyvSergl 586
DB 681 CTACTCTGCTCAGCCATCTACGAGAGGAGTGCAGTACTTGTCTTACAGCTGCAGCCA 740
QY 586 nleuProThrCyvsglyAspGlnValYrSerValaArgArgGlnlyvProserAspArg 606
DB 741 GCTGCCCACTTGGAGAGCAAGTCAATTCGTGGCCAGGCGGCAAGCAAGTACAG 800
QY 606 gAlaAspSerArgArgSerTyrHiasgluInserProPhegluYvGlnPheylvArgArg 626
DB 801 AGCTGACTCGCGCGGAGCTGGCATMAAGAGGCCCTTTGAAACAGATTTAAAGCAG 860
QY 626 gSerCyvsglnMetgluPheglYgluserIleMetserGluuAspArgSergluGlu 646
DB 861 AAGCTGCCAAATGAAATTTGAGAGAGCATCATGTAGAGAAAGAGTACGGGAAAGAGCT 920
QY 646 uGlyvValaglYserGlnserSerPheSerGlySerMetgluIleIlegluValSer 665
DB 921 GGGGAAAGTGGGAGTCACTTAAGCTTTTGGGAGCAGATGAATCATTTGAGGTCTCC 978

RESULT 28
AK098310 1916 bp mRNA linear PRI 15-JUL-2002
LOCUS AK098310
DEFINITION Homo sapiens cDNA FLJ40991 fis. clone UTRU0015198, weakly similar
to DUAL SPECIFICITY PROTEIN PHOSPHATASE 8 (BC 3.1.3.48).
ACCESSION AK098310

VERSION AK098310.1 GI:21758299
 KEYWORDS oligo capping; fis (full insert sequence).
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
 1 Kawakami, B., Sugiyama, A., Takemoto, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagahtsuna, M., Murakawa, K., Kanemori, K., Takahashi-Fujii, A., Oshina, A., Suzuki, Y., Sugano, S., Nagahara, K., Masuno, Y., Nagai, K., and Isogai, T.
 NEDO human cDNA sequencing project
 Unpublished
 2 (bases 1 to 1916)
 Isogai, T. and Yamamoto, J.
 Direct Submission
 Submitted (04-JUL-2002) Takao Isogai, FUJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan
 (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
 NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

FEATURES
 source
 1. 1916
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="UTERU2015198"
 /issue_type="uterus"
 /clone_id="UTERU2"
 /note="Cloning vector: pME18SFL3"

ORIGIN
 Alignment Scores:
 Pred. No.: 1,42e-164 Length: 1916
 Score: 178.00 Matches: 278
 Percent Similarity: 99.29% Conservative: 0
 Best Local Similarity: 99.29% Mismatches: 1
 Query Match: 26.77% Indels: 2
 DB: 9 Gaps: 0

US-10-029-345A-109 (1-665) x AK098310 (1-1916)

QY 387 AlaAspArgLeuGluAspSerAsnLysLeuLysArgSerPheSerLeuAspIleLeuSer 406
 DB 142 GCGAGCAGCTCGAAGACACGCAATTAAGCTTCAGAGCTTCTCTCTGATATCAATCA 201
 QY 407 ValSerTyrSerAlaSerMetAlaAspSerLeuHISGlyPheSerSerSerGluAspAla 426
 DB 202 GTTTCATATTCAGCCGACATGCGACATCTTACATGCTTCTCTCATCAGAAATGCT 261
 QY 421 LeuGluTyrTyrTyrProSerThrThrLeuAspGlyThrAsnLysLeuGlnPheSer 446
 DB 263 TTGGATATCTACAAACCTTCACATCTCTGATGGAGCAACAACTATGACGATTCCTCC 321
 QY 447 ProValGlnGluLeuSerGluGlnThrProGluThrProAspLysGlnGluAlaSer 466
 DB 322 CCTGTTACGAACTATTCGAGACAGACTCCCGAAACCAAGTCTGATTAAGAGGAGACCGAC 381
 QY 467 IlePolyLeuLeuGlnThrAlaArgProSerAspSerGlnSerLysArgLeuHisSer 486
 DB 382 ATCCCAAGAAAGCTCAGACCGCCAGGCTTTCAGACAGCCAGACAGAGGATGATTCG 441
 QY 487 ValArgThrSerSerSerGlyThrAlaGlnArgSerLeuSerProLeuHisArgSer 506

DB 442 GTGAGAACGACGACGAGTGGACCGCCAGAGATGCTCTTATATCTCACTGCATCGAAGT 501
 QY 507 GlySerValGluAspAsnTyrHisThrSerPheLeuPheGlyLeuSerThr-SerGlnG 526
 DB 502 GGGAGCGTGGAGGAGAAATTAACACACAGCTTCCTTTCCGGCTTTCAT-CAGCCAGCA 560
 QY 526 HisLeuThrLysSerAlaGlyLeuGlyLeuLysGlyTyrPheSerAspIleLeuAlaP 546
 DB 561 GCACCTTCAGAAAGTCTGCTGGCTGGCTGGCTTAAAGGCTGGACCTGCATATTTGGCCCC 620
 QY 546 GlnThrSerThrProSerLeuThrSerSerTyrTyrPheAlaThrGlnSerSerHisph 566
 DB 621 CCAGACCTTACCCCTTCCTTCGACACGACGCTGATATTTGGCAGACAGATCCTCACACT 680
 QY 566 eTyrSerAlaSerAlaIleTyrGlyGlySerAlaSerTyrSerAlaTyrSerCysSerG 586
 DB 681 CTACTCTCTCCTAGCATCTACGAGGAGGAGGAGTCCAGTACTCTGCTCAGCTGACGCA 740
 QY 586 nLeuProThrCysGlyAspGlnValTyrSerValArgArgGlnLysProSerAspArg 606
 DB 741 GCTGCCCACTTCGCGAGACCAAGTATTTCTGTGGCAGCGCGCAGAACCCAGAGCAG 800
 QY 606 gAlaAspSerArgArgSerTyrPheIleGlnGluSerProPheGlnLysGlnPheLysArg 626
 DB 801 AGCTGACTCGCGCGAGCTGGCATGAGAGAGCCCTTTGAAAGCAGTTTAAACGCAG 860
 QY 626 gSerCysGlnMetGluPheGlyGlnSerTyrMetSerGluAsnArgSerArgGlnGluLe 646
 DB 861 AAGCTGCCAAATGGAATTTGAGAGAGCATATGTCAGAGAAACAGTCAACGAGGAGACT 920
 QY 646 uGlyValGlySerGlnSerSerPheSerGlySerMetGluIleIleGluValSer 665
 DB 921 GGGGAAATGGGACGATGATGCTTTCGGGACGATGAGAAATCATTAAGTCTCC 978

RESULT 29
 AC097818 244605 bp DNA linear HTG 10-MAY-2003
 LOCUS Rattus norvegicus clone CH230-52K22, WORKING DRAFT SEQUENCE.
 AC097818
 DEFINITION AC097818.7 GI:30520485
 ACCESSION AC097818
 VERSION
 KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE
 1 (bases 1 to 244605)
 Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Albrooke, S., Amin, A., Angiano, D., Anyalbechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F., Biwalto, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Bunay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Centner, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Dudzin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, A., Falls, T., Fan, G., Fernandez, B., Flagg, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W., Gunaratne, P., Haaland, W., Hamli, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogue, M., Hollins, B., Howell, S., Huily, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpach, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, U., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Louised, H., Lozad, R. J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,

Manqum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mawlin, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
 Mijavajevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Nandiasa, M., Murphy, M., Nair, L.,
 Nantervy, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
 Nwakoelameh, O., Okunolu, G., Olarunbasogun, A., Pal, S., Parks, K.,
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C.,
 Plopper, F., Polidexter, A., Popovic, D., Primus, E., Pu, L.-L.,
 Puazo, M., Quiriz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ritz, S., J.,
 Sanders, M., Saverly, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smajls, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
 Steidle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umanu, K.,
 Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, P.,
 Williams, G., Willison, R., Wleczyk, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
 Weinstock, G. and Gibbs, R.A.

Unpublished
 Direct Submission
 2 (bases 1 to 244605)
 Worley, K.C.
 Direct Submission
 Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 244605)
 Rat Genome Sequencing Consortium.
 Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA

On May 10, 2003 this sequence version replaced gi:24819653.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by fixed gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu

----- Project Information
 Center project name: GFRK
 Center clone name: CH230-52K22
 ----- Summary Statistics
 Assembly program: Atlas 3.0:
 Consensus quality: 234323 bases at least Q40
 Consensus quality: 235890 bases at least Q30
 Consensus quality: 236794 bases at least Q20
 Estimated insert size: 241565; sum-of-contigs estimation
 Quality coverage: 8x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have

* provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 1 244605: contig of 244605 bp in length.
 Location/Qualifiers
 1..244605
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-52K22"
 complement(165..989)
 /note="clone boundary
 clone end: Sp6
 site: EcoRI
 end sequence: BH316904"
 23405..124172
 /note="clone boundary
 clone end: T7
 site: EcoRI
 end sequence: BH316901"
 238775..239794
 /note="wgs_end_extension
 clone end: T7"
 243232..244605
 /note="wgs_end_extension
 clone_end:T7"

ORIGIN
 Alignment Scores:
 Pred. No.: 1..2e-57 Length: 244605
 Score: 71.00 Matches: 71
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 10.68% Indels: 0
 DB: 2 Gaps: 0

US-10-029-345a-109 (1-665) x AC097818 (1-244605)
 QY 7 GlyThrGlnIleValThrGluArgLeuValAlaLeuLeuGluSerGlyThrGluVal 26
 Db 123608 GGAACCAATATGTTACTAGAGGTTGTGGCTCTGTGGAAAGCGCACCGAAGTG 123667
 QY 27 LeuLeuIleASPSeRarGpRopheValGluTYrAantThSerHisIleLeuGluAlaIle 46
 Db 123668 CTGCTAATGACAGCGGCGCATTGTGAGTACAAATACGTCACATCTCGAAGCCATC 123727
 QY 47 AsnIleAsnCySerIysLeuMetIysAArgArgLeuGlnIleAspIysValLeuIleThr 66
 Db 123728 AATATCACTGCTCCAACTGATGAGCAAGGCTCCAAACAAAGATTATTAACA 123787
 QY 67 GluLeuIleGlnHisSerAlaIysVal 77
 Db 123788 GAACTCATCCAGCATTCCTCAAAACATAAGTA 123820

RESULT 30
 AB052157
 LOCUS 4874 bp mRNA linear ROD 18-OCT-2001
 DEFINITION Mus musculus MKP-7 mRNA for MAP kinase phosphatase-7, complete cds.
 AB052157
 ACCESSION AB052157
 VERSION AB052157.1 GI:13990988
 KEYWORDS
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 1 Masuda, K., Shima, H., Watanabe, M. and Kikuchi, K.
 MKP-7, a novel mitogen-activated protein kinase phosphatase,
 functions as a shuttle protein
 J. Biol. Chem. 276 (42), 39002-39011 (2001)
 JOURNAL MEDLINE 21486429
 PUBMED 11489891

REMARK

Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
CDNA Library Preparation: M. Bento Soares, University of Iowa
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLB)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome Center, Stanford University School of Medicine, Stanford, CA 94305
Web site: <http://www-shgc.stanford.edu>
Contact: (Dickson, Mark) mcdepaxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers, R. M.

FEATURES

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLB at: <http://image.lnl.gov>
Series: IRAK Plate: 125 Row: 1 Column: 11
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 18485489.
Location/Qualifiers

source

1. 4943

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6"

/db_xref="taxon:10090"

/clone="MGC:66821 IMAGE:6409944"

/cissue_type="Brain, enriched mouse brain 12.5dp"

/clone_lib="NIH BMAP_F00"

/lab_host="DH10B"

/note="Vector: pYX-ASC"

1. 4943

/gene="Dusp16"

/note="synonyms: MKP-7, MKP7, D6Ertd23e"

/db_xref="Locustid:70686"

/db_xref="GI:37748379"

501. 2483

/codon_start=1

/product="Dusp16 protein"

/protein_id="AAH59232.1"

/db_xref="GI:37748379"

/db_xref="Locustid:70686"

/translation="MAHEMIGTQIVTESVALLSEGTKEVLLIDSPREVTNTHLIE
AININSKLMKRLQDDKYLITELIHSKHKDIDCNQVTVDDSSDDVSLSDC
PLVLGLKLEERSNVHLALGGFAEFSRCPGLCEKSTLVPCLISQCLPVANIGPT
RIPLNLYLGGORDVNLKMDQNGIVYNASNTCPKPIEPHSFLRPVNDPCEK
ILPMLDKSVDFIEKAKASNGCVLIHCLAGISRSATYAIAYIMKMDMSIDEAYRFXE
KRPTTSPNFMGOLMDYEKTINNOTGMSGPKSLKLHLDKRSEVPVPAASAGRWKSA
LSLSPPCANSTSAASGQRLVHRAVSPLPSLLEDSPLYOALSGLDLSSEKLEDTKL
KKSFSLDIKSVISASMAASLHGFSSSEALDYKPSATLIDGNKLQCFSPVDVSRQS
PETSPPKEEHAIEKQPPRPSPESQVTRALSVYKTSAGSTORPFPPLHRSQSVENY
HTWFLGLSTSQOHLTKSAGLGLKGMHSDILAPQSSAPLTSWYFATSPSHLYSAA
IKGNSYSAVSCGLPTCSDOIYSVRROKPLDRADSRSMHESPPFKQFRRSCQ
MERGESIMENRRELKGVSGSSFSGSMELIEVS"

549. 899

/note="RHOD; Region: Rhodanese Homology Domain"

/db_xref="CDD:smat00450"

972. 1388

/note="DSPC; Region: Dual specificity phosphatase.
catalytic domain. Ser/Thr and Tyr protein phosphatases.
The enzyme's tertiary fold is highly similar to that of
lysine-specific phosphatases, except for a 'recognition'
region"

/db_xref="CDD:pfam00782"

ORIGIN

Alignment Scores:

Pred. No.: 4,27e-55 Length: 4943
Score: 67.00 Matches: 266
Percent Similarity: 95.00% Conservative: 0
Best Local Similarity: 95.00% Mismatches: 7
Query Match: 10.08% Indels: 14

DB: 10 Gaps: 0
US-10-029-345a-109 (1-665) x BC059232 (1-4943)

QY 15 LeuValAlaLeuLeuGluSerGlyThrGluValLeuLeuLeuLeuSerArgProPhe 34
DB 543 TTGGTGGCTCTGCTGGAAAGTGAACGAAAGATGCTCTGAATGATGACCAATTT 602
QY 35 ValGluTyrAsnThrSerHisLeuGluAlaLeuAsnLeuAsnCySerLeuMet 54
DB 603 GGGAAATACAAATACGCTCATTGTTGGAAGCCATTAAATCAACTGCTCAACGAA 662
QY 55 LysArgArgLeuGlnGlnAspValLeuLeuThrGluLeuLeuGlnHisSerAlaLys 74
DB 663 AACGAGGTTCACACGACGACCAAGATTAATTAAGAACTTATCCAACTTCTGCAAG 722
QY 75 HisLysValAspIleAspCysSerGlnLysValValTyrAspGlnSerSerGlna 94
DB 723 CATTAAGTTGACATTGACGAA-TCAGAAAGGTGATGTTATGATCAAGTCCCAAG 781
QY 94 PValAla-SerLeuSerSerArgCysPheLeuThrValLeuLeuGlyLysLeuGluLys- 113
DB 782 TGTGG-TTCTGTCTGTCAGACTGCTTCTCACTGCTTCTGGGTAAACTGGAGAG- 839
QY 114 SerPheAsnSerValHisLeuLeuAlaGlyGlyPheAlaGluPheSerArgCysPhePro 133
DB 840 AGCTTCACTGTCCTCACTGCTGCAAGTGGCTTTCGAGTTCCTGCTGCTGCTTCCCT 899
QY 134 GlyLeuCyGluGlyLysSerThrLeuValProThrCysIleSerGlnProCysLeuPro 153
DB 900 GGGCTGTGGAAGAAAGTCACTGATGCTCACTGATGCTCACTGATGCTCACTGCTGCT 959
QY 154 ValAlaAsnIleGlyProThrArgIleLeuProAsnLeuThrLeuGlyCysGlnaAsp 173
DB 960 GTTGCACAACTGGGCACTGCAATTTCTCCAACTTCTGCTGCTGCTGCTGCTGCTGCT 1019
QY 174 ValLeuAsnLysGlu-LeuIle-GlnGlnAsnGlyIleGlyTyrValLeuAsnAlaSer 193
DB 1020 GTCTTCAACAGCA-CTTAT-GCAACAGATGGATTTGGCTTATGTTTAAATGCCAGCA 1077
QY 193 Yr-ThrCysProLysProAsnPheIleProGlnSerHisPheLeuArgValProValAsn 212
DB 1078 A-TACTGTCCAAAGCTGACTTCACTGATGCTCACTGCTGCTGCTGCTGCTGCTGCT 1136
QY 213 AspSerPheCysGluLysIleLeuProThrPheLeuAspLysSerValAspPheIleGlyLys 232
DB 1137 GACAGCTTTTGGAGAAATCTTACCATGTTGGACAACTGCTGCTGCTGCTGCTGCTGCT 1196
QY 233 AlaLysAlaSerAsnGlyCysVal-LeuValHisCysLeuAlaGlyIleSerArgSerAl 252
DB 1197 GCAAAAGCTCCAAAGCTGCTGCTGCTTATC-CAGTCTTACGCTGGATCTTCTGCTCCG 1255
QY 252 ThrIleAlaIleAlaTyrIleMetLysArgMetSerLeuAspGluAlaTyrArg 272
DB 1256 CACTATGCTTATGCTTCACTGATGATGAGAGATGACAGTCTCTGATGAGAGCTTACAG 1315
QY 272 GpHeValLysGluLysArgProThrIleSerProAsnPheAsnPhe 287
DB 1316 ATTGTGAAGAAAAAGACCTACTATCTCTCGAATTTTAATTTT 1361

RESULT 32
BC057321 4975 bp mRNA linear ROD 08-OCT-2003
LOCUS BC057321
DEFINITION Mus musculus dual specificity phosphatase 16, mRNA (CDNA clone
MGC:66822 IMAGE:6830600), complete cde.
ACCESSION BC057321
VERSION BC057321.1 GI:34980886
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 4975)

AUTHORS Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G., Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schlier,G.D., Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Datchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L., Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L., Schenck,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S., Cherniack,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mullaly,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S., Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W., Villalón,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodriguez,S., Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y., Souffarz,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D., Butterfield,Y.S., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Dickson,M.C., Krzywinski,M.I., Skalska,U., Smallus,D.E., Scherch,A., Schein,U.E., Jones,S.J., and Marra,M.A.

TITLE Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

MEDLINE 22388257

PUBMED 12477932

REFERENCE 2 (bases 1 to 4975)

AUTHORS Strausberg,R.

TITLE Direct Submision

JOURNAL Submitted (02-SEP-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk
Email: cgabs-r@mail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library Preparation: M. Bento Soares, University of Iowa
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@cgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susana Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Gai, Leticia Heiao, Martin Krzywinski, Keta Kutsche, Oliver Lee, Soo Sen Lee, Victor Ling, Carrie Mathewson, Candice McEady, Steven Ness, Pawan Pandoh, Anna-Lisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smallus, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranda Tsai, Natassja van den Bosch, Jill Vardy, George Yang, Scott Zuyderduyn, Marco Marra.

FEATURES

source

1..4975

location/Qualifiers

1..4975

organism="Mus musculus"

mol_type="mRNA"

strain="C57BL/6"

db_xref="taxon:10090"

clone="MGC:66822 IMAGE:6830600"

tissue_type="Brain, enriched mouse brain 12.5dp"

clone_lib="NIH_BMAP_F00"

lab_host="DH10B"

note="Vector: pYX-ASC"

1..4975

gene="Dusp16"

note="synonyms: MKP-7, MKP7, D6Ertc213e"

db_xref="locusID:70686"

db_xref="MGI:1917936"

565..2547

codon_start=1

product="Dusp16 protein"

protein_id="AAH57321.1"

/db_xref="GI:3498087"

/db_xref="locusID:70686"

/translation="MAHMITGQIVTSBVLALLSGETEKVLLIDSRPEVYNTSHILE
ATINCSKLMKRRLQDDKVLITELIQHAKRKVIDCQKVVYDQSSQDVSSSDC
FLTVLGLKERSFNSVNLIAAGFAFSCFPGLCGKSTLVPTCISQCLPVANIGPT
RLTVALGECORVDLMQONGGYVLANSNTCPKDFIPESHFLRPVNDCEK
ILPMTDSEVDFLEKXANSGCVLITCLAGISRAITAIYIMKMDMLDAEYPRKE
KRPTSPNPNMGOIMDYEXTINNOTGSGSKTKLILHDKPSRPVAAEBCGKSA
LSLSPCANSTSEAGQVLHPAPVPSLQPSLBDSPVLQALSGLQSSKEQSDTKL
KRSEFLDKSVSYSAKSMASLHGESSEALDYVPASVATLDGNTNKLQPSPVQESQ
PETSFDKKEAHIPKOPPPRPSQVTVTLHSVRTSSGSGTORPFRSPILRGSVEDNY
HTNPLFGISTQOHLTKSAGLGLKGMHDIAPQSSASPTSSWYFAEPHLYASA
ITGNSYSYAVSCGOLPFCSPDQIVSVRRORPDTBRADSRBSMHESPREKQPKRSCQ
MEFGSIENSEKRSRELEIKVSGSGSFSGSMETIYVS"

613..963

/note="RHOD; Region: Rhodanese Homology Domain"

/db_xref="CDD:smat0450"

1036..1452

/note="RSPc; Region: Dual specificity phosphatase, catalytic domain. Ser/Thr and Tyr protein phosphatase. The enzyme's tertiary fold is highly similar to that of tyrosine-specific phosphatases, except for a 'recognition region"

/db_xref="CDD:pfam00782"

ORIGIN

Alignment Scores:

Pred. No.: 4,29e-55 Length: 4975

Score: 67.00 Matches: 266

Percent Similarity: 95.00% Conservative: 0

Best Local Similarity: 95.00% Mismatches: 7

Query Match: 10.08% Indels: 14

DB: Gaps: 0

US-10-029-345A-109 (1-665) x BC057321 (1-4975)

QY 15 LeuValAlaLeuLeuGluSerGlyThrGluValLeuLeuIleAspSerArgProPhe 34

Db TTGGTGGCTCTGCTGGAAAGGGAACGGAAGAAAGTCTGCTAATGATGCGGACCATTT 666

QY 35 ValGluTyrAsnThrSerHisIleLeuGluAlaIleAsnIleAsnGlySerIysLeuMet 54

Db GTGGAATACATATACGTCATCATTTTGGAAAGCCATTAATATCAACTGCTCCAAACTGATG 726

QY 667 GTGGAATACATATACGTCATCATTTTGGAAAGCCATTAATATCAACTGCTCCAAACTGATG 726

QY 55 LysArgArgLeuGlnGlnAspIleValLeuIlePheGluLeuIleGlnHisSerAlaIys 74

Db AAGCGAAGTTGGCAACGAGCAAGATTAATATCAAGACTTATCCAAACTTCTGCAAG 786

QY 75 HisLysValAspIleAspCysSer-GlnLysValIleValItyrAspGlnSerSerGlnAs 94

Db CATAAAGTTGACATTGACTGCAATCAAGAGTGTAGTTATGATCAAAAGTCCCAAGA 845

QY 94 PValAla-SerLeuSerSerAspCysPheLeuThrValLeuLeuGlyLysLeuGluLys- 113

Db TGTGTGG-TTCTCTGTCCTGACAGCTGCTTCTCACTGATCTTGGGTAACCTGGAGAG-G 903

QY 114 SerPheAsnSerValHisLeuLeuAlaGlyGlyPheAlaGluPheSerArgCysPhePro 133

Db AGCTTCAACTCTGTCACCTGCTGACAGGAGGAGCTTGTAGTCTCTGCTTGTCTCT 963

QY 134 GlyLeuCysGluGlyLysSerThrLeuValProThrCysIleSerGlnProCysLeuPro 153

Db GGCCTCTGTGAAGGAAGTCCACTGATGCTTACCTGATCTGATCTGAGCTTGTCTTACT 1023

QY 154 ValAlaAsnIleGlyProThrArgGlyLeuLeuProAsnLeuTyrLeuGlyCysGlnArgAsp 173

Db GTTGGCAACATTGGGCGCACTCGAATCTTCCAAATCTCATCTTGTGCTGGCAGCAGAT 1083

QY 174 ValLeuAsnLysGlu-LeuIle-GlnGlnAsnGlyIleGlyTyrValIleAsnAlaSerT 193

Db GTCTCTCAACAGGA-CTTGAT-GCAACAGATGAGATGTGCTTATGTGTTAAAGCCAGCA 1141

QY 193 Yr-ThrCysProLysProAspPheIleProGluSerHisPheLeuArgValProValAsn 212

```

Db      1142 A-TACCTGTCACAAAGCTGACTCATATACGAAATTCACCTCTCGAGAGCTGTGAAT 1200
      |||
Qy      213 AAGSerPheGagGluYsIleLeuPProTPrLeuAAspYsSerAlaAAspIleGluYs 232
      |||
Db      1201 GACAGCTTTTGTGAGAAATCCATACATGTTGGACAAGCTGTGATTTGATTCATTGGAGAA 1260
      |||
Qy      233 AAlaYsAlaSerAngIysCysVal-LeuValHisCysLeuAlaGlyTlSerArgSerAl 252
      |||
Db      1261 GCAAAAGCTTCACATGCTGTGCTTATC-CACCTGTTAGCTGGATCTCTGCTCCGC 1319
      |||
Qy      252 aThrlleAlaIleAlaTyrIleMetIysArgMetAAspMetSerLeuAAspIuAlaTyr 272
      |||
Db      1320 CACTATTGCTGATTCCTACATCATGAAGAGATGAGATCATCTCTAGATAGGCTTACAG 1379
      |||
Qy      272 gPheValIysGluYsAlaArgProThrIleSerProAAspAAspHe 287
      |||
Db      1380 ATTGTGAAAGAAAAAGAACTACTATATCTCCGAATTTAATTTT 1425
      |||

RESULT 33
AC133722 242590 bp DNA linear HTG 20-NOV-2002
LOCUS Rattus norvegicus clone CH230-240110, *** SEQUENCING IN PROGRESS
DEFINITION
***, 3 unordered pieces.
AC133722
AC133722.2 GI:25138890
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus
1 (bases 1 to 242590)
Muzny,D.,Marle,M.,Metzker,M.,Lee,S.,Abramson,S.,Adams,C.,Alber,J.,
Allen,C.,Allen,H.,Alsbrooks,S.,Amin,A.,Anguiano,D.,
Anyalebechi,V.,Aoyagi,A.,Ayodeji,M.,Baca,B.,Baden,H.,
Baldwin,D.,Bandaranecke,D.,Barber,M.,Barnstead,M.,Benahmed,F.,
Biswal,K.,Blair,J.,Blankenburg,K.,Blyth,P.,Brown,M.,
Bryant,N.,Buhay,C.,Burch,P.,Burrell,K.,Calderon,E.,
Cardenas,V.,Carter,K.,Cavazos,I.,Ceasar,H.,Center,A.,
Chacko,J.,Chavez,D.,Chen,G.,Chen,R.,Chen,Y.,Chen,Z.,Chu,J.,
Cleveland,C.,Cockrell,R.,Cox,C.,Coyle,M.,Cree,A.,D'Souza,L.,
Davila,M.L.,Davis,C.,Davy-Carroll,L.,De Anda,C.,Dederich,D.,
Delgado,O.,Denson,S.,Deramo,C.,Ding,Y.,Dinh,H.,Diya,K.,
Draper,H.,Dugan-Rocha,S.,Dunn,A.,Durbid,K.,Duvall,B.,Eaves,K.,
Egan,A.,Escotto,M.,Eugene,C.,Evans,C.A.,Falls,T.,Fan,G.,
Fernandez,S.,Finley,M.,Flagg,N.,Forbes,L.,Foster,M.,Foster,P.,
Fraser,C.M.,Gabisi,A.,Ganta,R.,Garcia,A.,Garner,T.,Garza,M.,
Gebregorgis,E.,Geer,K.,Gill,R.,Grady,M.,Guerra,W.,Guevara,W.,
Gumaratne,P.,Haaland,W.,Hamil,C.,Hamilton,C.,Hamilton,K.,
Harvey,Y.,Havlik,P.,Hawes,A.,Henderson,N.,Hernandez,J.,
Hernandez,R.,Hines,S.,Hladun,S.L.,Hodgeson,A.,Hogues,M.,
Hollins,B.,Howells,S.,Hulys,S.,Hume,J.,Idlebird,D.,Jolivet,A.,
Jackson,L.,Jacob,L.,Jiang,H.,Johnson,B.,Johnson,R.,Jolivet,A.,
Karpathy,S.,Kelly,S.,Kelly,S.,Khan,Z.,King,L.,Kovar,C.,
Kowalski,C.,Kraft,C.L.,Lebow,H.,Levan,J.,Lewis,L.,Li,Z.,Liu,J.,
Liu,J.,Liu,W.,Liu,Y.,London,P.,Longacre,S.,Lopez,Z.,
Lorenshewer,L.,Loulesed,R.J.,Lozado,R.J.,Lu,X.,Ma,J.,
Maheshwari,M.,Mahindaratne,M.,Mahmoud,M.,Mallory,K.,Mangum,A.,
Mangum,B.,Mapua,P.,Martin,K.,Martin,R.,Martinez,E.,
Mawhinney,S.,McLeod,M.P.,McNeill,T.Z.,Meenen,E.,
Mlisaavljajic,A.,Miner,G.,Minja,E.,Montemayor,J.,Moore,S.,
Morgan,M.,Morris,K.,Morris,S.,Munidas,M.,Murphy,M.,Nair,L.,
Nankervis,C.,Neal,D.,Newton,N.,Nguyen,N.,Norris,S.,
Nwaokemele,O.,Okwuonu,G.,Olarnpunagoon,A.,Pal,S.,Parks,K.,
Pasternak,S.,Paul,H.,Perez,A.,Perez,L.,Pfankoch,C.,
Plomper,F.,Polindexter,A.,Popovic,D.,Primus,B.,Py,L.-L.,
Puzo,M.,Quiroz,J.,Rachlin,E.,Reeves,K.,Regier,M.A.,Reigh,R.,
Reilly,B.,Reilly,M.,Ren,Y.,Reuter,M.,Richards,S.,Riggs,F.,
Rives,C.,Rodkey,T.,Rojas,A.,Rose,M.,Rose,R.,Ruiz,S.J.,
Sanders,W.,Savery,G.,Scherer,S.,Scott,G.,Shatsman,S.,Shen,H.,
Shetty,J.,Shvartbeyn,A.,Sisson,I.,Sitter,C.D.,Smajs,D.,
Sneed,A.,Sodergren,E.,Song,X.-Z.,Sorelle,R.,Soes,J.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: KClE
Center clone name: CH230-240110
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 216426 bases at least Q40
Consensus quality: 220448 bases at least Q30
Consensus quality: 223410 bases at least Q20
Estimated insert size: 224864; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 240024: contig of 240024 bp in length
240025 240124: gap of unknown length
240125 241173: contig of 1049 bp in length
241174 241273: gap of unknown length
241274 242590: contig of 1317 bp in length.
Location/Qualifiers
1. 242590
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-240110"

```



```

misc_feature      1..1094
                  /note="wgs end extension
                  clone_end:Sp6"
misc_feature      7392..8314
                  /note="clone boundary
                  clone_end:Sp6
                  site:
end sequence:BZ106067"
misc_feature      complement(235121..235967)
                  /note="clone boundary
                  clone_end:T7
                  site:
end sequence:BZ106066"
misc_feature      236448..237947
                  /note="wgs end extension
                  clone_end:T7"
misc_feature      237998..240024
                  /note="wgs end extension
                  clone_end:T7"

ORIGIN
Alignment Scores:
Pred. No.:      9,83e-53      Length:      242590
Score:          66.00      Matches:      66
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    9.92%      Indels:      0
DB:             2      Gaps:      0

US-10-029-345A-109 (1-665) x AC133722 (1-242590)

QY      595 TysrValArArGrArGInlySPoSsrArpArAlaSPsrArGrArGrserTrpHis 614
Db      12283 TATTCGTTCGTAAGGGCGGCAAGCTTATGACAGCTGACTCAGCGGCGAGCTGGCAAT 12342

QY      615 GIUGIusePrPheGluylusGInPheLyArGrArSerCyGslmGluPheGluylu 634
Db      12343 GAAAGAGAGCCCTTTCAGAAAGCAGTTTAAACGAGAGAGCTGCCAATATGCAATTTGGAGAG 12402

QY      635 SerIleMetSerGIuAsnArGrSerArGrGluLeuGlylValGlySerGlnserSer 654
Db      12403 ACCATTATATCTCAGAAACAGCTCGAGGAAAGACGCGCAAGTGGGAGCCCAATCTTAC 12462

QY      655 PreSerGIySerMetGlu 660
Db      12463 TTTTCAGAAAGCATTGAG 12480

RESULT 34
AC128093/c.      250782 bp      DNA      linear      HTG 20-NOV-2002
LOCUS      AC128093
DEFINITION      Rattus norvegicus clone CH230-84N7, *** SEQUENCING IN PROGRESS ***.
ACCESSION      AC128093
VERSION      AC128093.3 GI:25137709
KEYWORDS      HTG; HTGS PHASE2; HTGS DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM      Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 250782)
Muzny,D,Marie., Metzger,M, Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Altsbrooks,S., Amin,A., Anguitano,D.,
Anyalelechi,V., Aoyagi,A., Ayodeji,M., Baca,B., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Devila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,

```

```

REFERENCE
AUTHORS      Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregorjis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,M.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Huylk,S., Hume,J., Idlibird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Louisedge,H., Losado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindaratne,M., Mahmood,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,S., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munday,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwackemeah,O., Okwoum,G., Olarpunagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfankoch,C.,
Plapper,F., Poindexter,A., Popovic,D., Prims,E., Pu,L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Taber,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Uemari,K.,
Valas,R., Vera,V., Villalana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., Wiley,K.,
Williams,G., Willson,R., Wiczzyk,R., Wooden,H., Wright,C.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,U., Zhou,X., Zhao,S., Dunn,D., von
Weintraubern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 250782)
Wortley,K.C.
Direct Submission
Submitted (19-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 250782)
Baylor College of Medicine
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 20, 2002 this sequence version replaced gi:23265205.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/r2/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GYAC
Center clone name: CH230-84N7
----- Summary Statistics
Assembly program: Phrap; version 0.990329

```


DEFINITION Mus musculus map kinase phosphatase-M B1 isoform mRNA, complete cds.
ACCESSION AF345953
VERSION AF345953.1 GI:13625396
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 4026)
AUTHORS Matsuguchi,T., Musikacharoen,T., Johnson,T.R., Kraft,A.S. and Yoshikai,Y.
TITLE A Novel MKP is an Important Negative Regulator of LPS-mediated JNK Activation in Macrophages
JOURNAL Mol. Cell. Biol. 20: 6999-7009 (2001)
REFERENCE Matsuguchi,T., Musikacharoen,T. and Yoshikai,Y.
AUTHORS Direct Submission
TITLE Submitted (07-FEB-2001) Medicine, Nagoya University, 65 Teinumaicho, Nagoya, Aichi 466-8550, Japan
FEATURES
source location/Qualifiers
1..4026
/organism="Mus musculus"
/mol_type="mRNA"
/strain="BALB/c"
/db_xref="taxon:10090"
/cell_type="macrophage"
558..1625
/note="MKP-M B1 isoform; protein phosphatase"
/codon_start=1
/product="map kinase phosphatase-M B1 isoform"
/protein_id="AAK35054.1"
/db_xref="GI:13625397"
/translation="MAHEMIGQIVTESIVALLSGTEKVLIDSRPFVNTSHILE
AINNCKIMKRRLOQDRVLTIELIGSAHKVIDDNQKYVVDSSQVGSLSDC
FLIVLGLKLEKSFNSVHLAAGFEAFRCPEGLCEKSTLVPTCTISQCLPVANIGPT
RILPLNLYGQCRDVLNKLQNGIGVLANASNTCPKDPFIPESHFLRPVNDSCFK
ILPWLIDSVFIEKAKASNGCVLIHCLAGISRSATIAIAYIMKMDMSIDEPYRQK
ILPWLIDSVFIEKAKASNGCVLIHCLAGISRSATIAIAYIMKMDMSIDEPYRQK
TPRADSRRTGMRAPLKSLSLNAEAKMNLERALCRRTPGRSMARMAASPAAMRS
SRSLKTSLLTLVLPVHK"

ORIGIN
Alignment Scores:
Pred. No.: 3,21e-52 Length: 4026
Score: 64.00 Matches: 64
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.62% Indels: 0
DB: 10 Gaps: 0
US-10-029-345a-109 (1-665) x AF345953 (1-4026)

QY 114 SerPheAsnSerValHisLeuLeuAlaGlyGlyPheAlaGlyPheSerArgCysPhePro 133
DB 897 AGCTTCAATCTGTCTCACCCTGCTGAGGCTTGCGAATCTCTGTTGTTCCCT 956
QY 134 GlyLeuCySGluGlyLysSerThrLeuValProThrCysIleSerGlnProCysLeuPro 153
DB 957 GGCCTCTGGAAGAAAGTCCACTGATCCTGATATCTCAGCGTTGCTTAACT 1016
QY 154 ValAlaAsnIleGlyProThrArgIleLeuProAsnLeuTyrLeuGlyCysGlnArgAsp 173
DB 1017 GTTGGAAACATTTGGGCAATCGAATCTTCCCAATCTCTATCTTGCGCAGAGAT 1076
QY 174 ValLeuAsnLys 177
DB 1077 GTCTCAACACAG 1088

RESULT 37
AF345951 4992 bp mRNA linear ROD 24-SEP-2001
LOCUS AF345951
DEFINITION Mus musculus map kinase phosphatase-M A1 isoform mRNA, complete cds.

ACCESSION AF345951
VERSION AF345951.1 GI:13625392
KEYWORDS
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 4992)
AUTHORS Matsuguchi,T., Musikacharoen,T., Johnson,T.R., Kraft,A.S. and Yoshikai,Y.
TITLE A Novel MKP is an Important Negative Regulator of LPS-mediated JNK Activation in Macrophages
JOURNAL Mol. Cell. Biol. 20: 6999-7009 (2001)
REFERENCE Matsuguchi,T., Musikacharoen,T. and Yoshikai,Y.
AUTHORS Direct Submission
TITLE Submitted (07-FEB-2001) Medicine, Nagoya University, 65 Teinumaicho, Nagoya, Aichi 466-8550, Japan
FEATURES
source location/Qualifiers
1..4992
/organism="Mus musculus"
/mol_type="mRNA"
/strain="BALB/c"
/db_xref="taxon:10090"
/cell_type="macrophage"
558..2591
/note="MKP-M A1 isoform; protein phosphatase"
/codon_start=1
/product="map kinase phosphatase-M A1 isoform"
/protein_id="AAK35052.1"
/db_xref="GI:13625393"
/translation="MAHEMIGQIVTESIVALLSGTEKVLIDSRPFVNTSHILE
AINNCKIMKRRLOQDRVLTIELIGSAHKVIDDNQKYVVDSSQVGSLSDC
FLIVLGLKLEKSFNSVHLAAGFEAFRCPEGLCEKSTLVPTCTISQCLPVANIGPT
RILPLNLYGQCRDVLNKLQNGIGVLANASNTCPKDPFIPESHFLRPVNDSCFK
ILPWLIDSVFIEKAKASNGCVLIHCLAGISRSATIAIAYIMKMDMSIDEPYRQK
ILPWLIDSVFIEKAKASNGCVLIHCLAGISRSATIAIAYIMKMDMSIDEPYRQK
LSPSPCANSTSESGORLHPASVSPILQSLIEDSPVQALSGQLSSEKLESTLU
KRSPLDIKYSVYSASMAAISHPSSSEALDYCKPSATLTGNTNLCQFSVQVEES
PEISPDKEAHIPKQPPRPSESQVTRLSHVRSGSTQRPFPFLASGVEEDY
HTNPLFGFHPQATPQCRAWLKGMHSDILAPQSAPSLTSSWYFTEPSHLYSAA
IYGNSSYSVASCGQLPTCSGDIYSVRRQPTDRAPSRRTGMRAPLKSLSLNAEAK
MNLERALCRRTPGRSMARMAASPAAMRSRSLKTSLLTLVLPVHK"

ORIGIN
Alignment Scores:
Pred. No.: 3,83e-52 Length: 4992
Score: 64.00 Matches: 93
Percent Similarity: 97.89% Conservative: 0
Best Local Similarity: 97.89% Mismatches: 1
Query Match: 9.62% Indels: 2
DB: 10 Gaps: 0
US-10-029-345a-109 (1-665) x AF345951 (1-4992)

QY 194 ThrCysProLysProAspPheIleProGluSerHisPheLeuArgValProValAsnAsp 213
DB 1137 ACCGTGCAAGCTGACTTCAATCTGAAATCTCACTTCTGGAAGCTGGAATGAC 1196
QY 214 SerPheCysGluLysIleLeuProTyrPheAspLysSerValAspPheIleGluLysAla 233
DB 1197 AGCTTTGTGAGAAATCTCAACATGAGTGGAGCAAGCTGTGATTCATTGAAAGCA 1256
QY 234 LysAlaSerAsnGlyCysVal-LeuValHisCysGluLeuAlaGlyTyrIleSerArgSerAla 253
DB 1257 AAAGCCTCCAAATGGCTGTGCTTATC-CACCTCTTAGCTGGGATCTCTGCTCCGAC 1315
QY 253 rIleAlaIleAlaTyrIleMetLysArgMetAspMetSerLeuAspGlnAlaTyrArgPh 273
DB 1316 TATTTGCTATTGCTCAATCATGAAAGAGATGACATGTCTCTGAGATGAGGCTTACAGAT 1375
QY 273 eValLysGluLysArgProThrIleSerProAsnAsnPhe 287

Db	1376	1GTGAAGAAAAAGCCACTATATCTCGATTATTTT	1418
RESULT 38			
LOCUS	AC126692		
DEFINITION	Mus musculus chromosome UNK clone RP23-246U18, WORKING DRAFT SEQUENCE, 9 unordered pieces.	201474 bp	DNA 11near HTG 12-OCT-2000
ACCESSION	AC126692		
VERSION	AC126692.2	GI:23915608	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.		
SOURCE	Mus musculus (house mouse)		
ORGANISM	Mus musculus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 201474)		
AUTHORS	McPherson, J.D. and Waterston, R.H.		
TITLE	The sequence of Mus musculus clone		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 201474)		
AUTHORS	McPherson, J.D. and Waterston, R.H.		
TITLE	Direct Submision		
JOURNAL	Submitted (08-JUL-2002) Genome Sequencing Center, 4444 Forest Park		
REFERENCE	3 (bases 1 to 201474)		
AUTHORS	McPherson, J.D. and Waterston, R.H.		
TITLE	Direct Submision		
JOURNAL	Submitted (12-OCT-2002) Genome Sequencing Center, 4444 Forest Park		
COMMENT	Pathway, St. Louis, MO 63108, USA		
	On Oct 12, 2002 this sequence version replaced gi:21703684.		

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: M BA0246318

----- Summary Statistics -----
Sequencing vector: M13, 0%
Sequencing vector: Plasmid, 100%
Chemistry: Dye-Primer ET, 0% of reads
Chemistry: Dye-terminator Big Dye, 100% of reads
Assembly program: Phrap, version 0.990319
Consensus quality: 197971 bases at least Q40
Consensus quality: 198628 bases at least Q30
Consensus quality: 199034 bases at least Q20
Insert size: 195000; agarose-fp
Insert size: 200674; sum-of-contigs
Quality coverage: 9.96 in Q20 bases; agarose-fp
Quality coverage: 9.99 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1
2368: contig of 2368 bp in length
2369
2468: gap of unknown length
2469
6036: contig of 3568 bp in length
6037
6136: gap of unknown length
6137
14367: contig of 8231 bp in length
14368
14667: gap of unknown length
14668
28520: contig of 14053 bp in length
28521
28620: gap of unknown length
28621
41185: contig of 12565 bp in length
41186
41285: gap of unknown length
41286
57620: contig of 16335 bp in length
57621
57720: gap of unknown length
74576: contig of 16856 bp in length
57721

```

[illegible]

REFERENCE
AUTHORS
TITLE
JOURNAL
2 (bases 1 to 201474)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (08-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 201474)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (12-OCT-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On Oct 12, 2002 this sequence version replaced gi:21703684.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WTJSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: M BA0246J18
----- Summary Statistics -----
Sequencing vector: MJ3, 0%
Sequencing vector: plasmid, 100%
Chemistry: Dye-terminator Big Dye, 100% of reads
Chemistry: Dye-terminator Big Dye, 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 197971 bases at least Q40
Consensus quality: 198628 bases at least Q30
Consensus quality: 199034 bases at least Q20
Insert size: 195000; agarose-fp
Quality coverage: 9.62 in Q20 bases; agarose-fp
Quality coverage: 9.99 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
2369 2368: contig of 2368 bp in length
2469 2468: gap of unknown length
6037 6036: contig of 3568 bp in length
6137 6136: gap of unknown length
14367 14367: contig of 8231 bp in length
14467 14467: gap of unknown length
28521 28520: contig of 14053 bp in length
28621 28620: gap of unknown length
41185 41185: contig of 12565 bp in length
41286 41285: gap of unknown length
57621 57620: contig of 16335 bp in length
57721 57720: gap of unknown length
74577 74576: contig of 16856 bp in length
74677 74676: gap of unknown length
125817 125817: contig of 51141 bp in length
125818 125917: gap of unknown length
125918 201474: contig of 75557 bp in length.

FEATURES
source
1..201474
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="10M"
/clone="RP23-246J18"
1..2368
/note="assembly_name:Contig14"
misc_feature
2469..6036
/note="assembly_name:Contig15"
6137..14367
/note="assembly_name:Contig16"
14468..28520
misc_feature

misc_feature
28621..41185
/note="assembly_name:Contig17"
misc_feature
41286..57620
/note="assembly_name:Contig18"
misc_feature
57721..74576
/note="assembly_name:Contig19"
misc_feature
74677..125817
/note="assembly_name:Contig20"
misc_feature
125918..201474
/note="assembly_name:Contig21"
misc_feature
/note="assembly_name:Contig22"
ORIGIN
Alignment Scores:
Pred. No.: 6,97e-48 Length: 201474
Score: 61.00 Matches: 61
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 9.17% Indels: 0
DB: 2 Gaps: 0
US-10-029-345A-109 (1-665) x AC126692 (1-201474)
QY 605 AsparGAlaAspSerArgSerTrpHisGluGluSerProPheGluValGlnPheLys 624
DB 16511 GACAGAGCTGACTGACGCGAGCTGCATGAAAGAGAGCCCTTTGAAAACAGTTTAAA 165052
QY 625 ArgArgSerCysGlnMetGluPheGlyGluSerIleMetSerGluAsnArgSerArgGlu 644
DB 165051 CGCAGAAAGCTCCCAATGGAATTGGAGAGACATTAATGTCGAGAACAGGTCCAGGGAG 164992
QY 645 GluLeuGlyValGlySerGlnSerSerPheSerGlySerMetGluIleGluVal 664
DB 164991 GAGCTGGGCAAGTGGGCGAGTCAGTTCCTTCGCGACCAAGAGATCATCGAGTGC 164932
QY 665 Ser 665
DB 164931 TCT 164929
RESULT 40
AC118035
LOCUS
DEFINITION Mus musculus clone RP23-174M4, WORKING DRAFT SEQUENCE, 7 unordered
pieces.
AC118035
VERSION AC118035.3 GI:30017966
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Baeten,V., Bloom,T., Boguslavsky,L.,
Bonkshalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Collange,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., DeArrellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Coyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Lacombe,K., Lamazares,R.,
Lander,T., Lehoczyk,J., Levine,R., Lindblad-Toh,K., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Marguis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Menues,L.,
Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunhahang,P., Pierse,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupack,R.,


```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="RZPD:DKFZp451D0719"
/db_xref="taxon:9606"
/map="12p13.2"
/clone="DKFZp451D0719"
/clone_type="human spinal cord"
/clone_id="451 (synonym: hlccl). Vector psport1; host
DH10B; sites NotI + SalI"
/dev_stages="adult"
5027

ORIGIN
polya_site

Alignment Scores:
Prd. No.: 2 01e-35 length: 5047
Score: 47.00 Matches: 47
Percent Similarity: 100.00% Conservative: 0
Beat Local Similarity: 100.00% Mismatches: 0
Query Match: 7.07% Indels: 0
DB: 9 Gaps: 0

US-10-029-345A-109 (1-665) x HSM804384 (1-5047)

QY 77 ValAspIleApcYssergIntlyValValValTYrAspGInserSergInaSpValAla 96
821 GTTGACATTGATTCGACGTCGAGAGCTTGCTTACGATCAAGCTCCCAAGATGTTGCC 880
DB 821 GTTGACATTGATTCGACGTCGAGAGCTTGCTTACGATCAAGCTCCCAAGATGTTGCC 880
QY 97 SerLeuSerSerApcYsPheLeuThrValIleuLeuGlyIylValLeuGlyIylSerPheAan 116
881 TCTCTCTCTTCACAGCTGTTTCTCAGCTGCTTGGGTAACTGAGAGAGAGCTTCAAC 940
DB 881 TCTCTCTCTTCACAGCTGTTTCTCAGCTGCTTGGGTAACTGAGAGAGAGCTTCAAC 940
QY 117 SerValHileuLeuAlaGly 123
941 TCTGTCACCTGCTTGCAGGT 961
DB 941 TCTGTCACCTGCTTGCAGGT 961

RESULT 44
AC097776 247148 bp DNA linear HTG 13-NOV-2002
LOCUS Rattus norvegicus clone CH230-94G4, *** SEQUENCING IN PROGRESS ***
DEFINITION 2 unordered pieces.
AC097776.5 GI:24942830
VERSION HTG: HTGS PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 247148)
REFERENCE
AUTHORS Muzny, D., Marie, Metzker, M., Lee, Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benhamed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blych, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A.,
Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Devila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
Delgado, O., Dengson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
Draper, H., Dungan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Georgiev, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W.,
Gunnarsson, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogue, M.,
Hollins, B., Howells, S., Huily, S., Hume, J., Idelbrid, D., Jackson, A.,
Jackson, B., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kowib, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,

```

```

Lorenshewa, L., Louised, H., Lozano, R.J., Lu, X., Ma, J.,
Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E.,
Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenen, B.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Mundasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwackemeke, O., Okunolu, G., Olarnunagoon, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C.,
Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., L.,
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorrelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villaseña, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczek, R., Woodson, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G., and Gibbs, R.A.
Direct Submision
Unpublished
2 (bases 1 to 247148)
REFERENCE
AUTHORS Worley, K.C.
TITLE Direct Submision
JOURNAL Submitted (23-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 247148)
REFERENCE
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submision
JOURNAL Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 13, 2002 this sequence version replaced gi:23269395.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GFPZ
Center clone name: CH230-94G4
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 208726 bases at least Q40
Consensus quality: 212396 bases at least Q30
Consensus quality: 214499 bases at least Q20
Estimated insert size: 210868; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is

```


* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 11276: contig of 11276 bp in length
* 11277 11376: gap of unknown length
* 11377 247148: contig of 235772 bp in length.

FEATURES
Location/Qualifiers
1..247148

misc_feature
1..1340
/note="wgs contig"
misc_feature
107191..108362
/note="wgs contig"
misc_feature
132629..136209
/note="wgs contig"
misc_feature
136341..137406
/note="wgs contig"

ORIGIN

Alignment Scores:

Pred. No.: 2,22e-14 Length: 247148
Score: 27.00 Matches: 27
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 4.06% Indels: 0
DB: 2 Gaps: 0

US-10-029-345A-109 (1-665) x AC097776 (1-247148)

QY 97 SerLeuSerSerAspGlyPheLeuThrValLeuLeuGlyLysLeuGlyLysSerPheLeu 116
Db 1077 TCTCTGCTCCTCAGACGCTTCTCTCCTGCTTGGGTAACTGGAGAGAGCTTCAAC 1136

QY 117 SerValHisLeuLeuAlaGly 123
Db 1137 TCTGTTCACTGCTGCTGAGGT 1157

RESULT 45

LOCUS AK116903 3540 bp mRNA linear INV 30-NOV-2002
DEFINITION Ciona intestinalis cDNA, clone:cieg018b07, full insert sequence.

ACCESSION AK116903.1 GI:23589623

VERSION FLI CDNA.

KEYWORDS Ciona intestinalis

SOURCE Ciona intestinalis

ORGANISM Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;

Phlebobranchia; Clonidae; Clona.

REFERENCE

AUTHORS

1 Satou, Y., Yamada, L., Mochizuki, Y., Takatori, N., Kawashima, T.,
Sasaki, A., Hamaguchi, M., Awazu, S., Yagi, K., Sasaki, Y.,
Nakayama, A., Ishikawa, H., Inaba, K. and Satoh, N.
A cDNA resource from the basal chordate Ciona intestinalis
Genesis 33 (4), 153-154 (2002)

JOURNAL MEDLINE
TITLE 2 (bases 1 to 3540)
REFERENCE PubMed
AUTHORS Satou, Y. and Satoh, N.
TITLE Direct Submission
JOURNAL Submitted (04-OCT-2002) Nori Satoh, Kyoto University, Department of
Zoology, Sakyo-ku, Kyoto, Kyoto 606-8502, Japan
(E-mail: satou@ascidian.zool.kyoto-u.ac.jp, Tel: 81-75-753-4095,
Fax: 81-75-705-1113)
Ciona intestinalis cDNA Project (URL:
http://ghost.zool.kyoto-u.ac.jp/indexr1.html).
Location/Qualifiers

COMMENT

FEATURES
source

1..3540
/organism="Ciona intestinalis"

/mol_type="mRNA"
/db_xref="taxon:7719"
/clone="cieg018b07"

ORIGIN

Alignment Scores:

Pred. No.: 5.73e-11 Length: 3540
Score: 22.00 Matches: 22
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.31% Indels: 0
DB: 3 Gaps: 0

US-10-029-345A-109 (1-665) x AK116903 (1-3540)

QY 270 AlaTyrArgPheValysGlyLysArgProThrIleSerProAsnPheAsnPheLeuGly 289
Db 842 GCTTACAGGTTCTGTAAGAGAAAAGACCAACCAATTTCTCCCACTTAAATTTCTCGGG 901

QY 290 GlnLeu 291
Db 902 CAGCTG 907

RESULT 46

LOCUS BC045110 2264 bp mRNA linear PRI 07-OCT-2003
DEFINITION Homo sapiens dual specificity phosphatase 8, mRNA (cDNA clone
MGC:52239 IMAGE:5478677), complete cds.

ACCESSION BC045110.1 GI:28277227

VERSION MGC.

KEYWORDS Homo sapiens (human)

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 2264)

REFERENCE

AUTHORS

1 Strausberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heide, F.,
Diachenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loggellano, N.A., Peters, G.J.,
Abramson, R.D., Mullahy, S.J., Bosak, S.A., Mewen, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hult, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahney, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smal, D.E.,
Scherer, A., Schein, J.E., Jones, S.J., and Marra, M.A.
Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

JOURNAL MEDLINE
TITLE 2 (bases 1 to 2264)
REFERENCE PubMed
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (27-JAN-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
Contact: MGC help desk
Email: gcgaps-remail.nih.gov
Tissue Procurement: ATCC

CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILUNI)
DNA Sequencing by: National Institutes of Health Intramural

Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgen@nri.nih.gov
Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blaesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dierich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S.-L., Karlins, E., Kong, P., Latic, P., Legaspi, R.,
Maduro, Q.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Stantitop, S., Thomas, P.J., Touchman, J.W.,
Tsugeon, C., Vogt, J.D., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/BLN at: <http://image.llnl.gov>
Series: IRAL Plate: 44 Row: 1 Column: 16
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 4758211.

FEATURES

source

1..2264
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:52239 IMAGE:5478677"
/issue_type="Brain, astrocytoma, grade IV"
/clone_lib="NIH MGC_98"
/lab_host="DH10B"
/note="Vector: pOTB7"
1..2264
/gene="DUSP8"
/note="synonyms: HB5, HVH-5, HVH8"
/db_xref="locusid:1850"
/db_xref="MIM:602038"
2..1879
/codon_start=1
/product="dual specificity phosphatase 8"
/protein_id="AAH45110.1"
/db_xref="GI:28277228"
/db_xref="LocusID:1850"
/translation="MAGDRLPRKVDKAKLASLRGPGGPIVDSRFSVEYNMHWL
SSVNTICSLKVRKLOQGVTTIAELIQPARSOVETBEPODVVVQSTRDASVLAAD
SFSLSLKLDGCFVALITGFAFFSCFPLCEGKPAALLPMSLSQPCIPVSG
LRIILPHLYIGSKVDLMTONGISIVLANSNGCPKDPICERMRPVDINDVC
EKLLPWLDSIERIDAKLSGCQVYHCLAGISRSRTTIAITIMKTWMSDDAIRFV
KORRPSISPNFNLGQLYERSLKULALQGPISGTEPPSPPAAGALPLRPLP
PTSESATGNAARBEGLSAGEPPAPPTPPATSAQQGLRGHLSDRLQDTNLRK
SFSLDIKSAVYASRRPDGPPDPGPAPLCKLDSGALGSSPSPDPAAPAR
PPRRRRPPRPPAGPARSPAHSLGLFGDAROTPRHGLSALSPGLRPGOPAGPCAW
APPLDSRPPSPDPGWCPSPEAGAGAGVLPAPFGAGAGPGGSGDLARRAARAP
RDARTGMPEEPAPETQFKRRSCOMEPEBGMVEGRANGELALGKQASFGSVEVTV
S"

gene

CDS

misc_feature
47..403
/note="RHOD; Region: Rhodanese Homology Domain"
/db_xref="CDD:smarr00450"
misc_feature
479..895
/note="DSPC; Region: Dual specificity phosphatase,
catalytic domain. Ser/Thr and Tyr protein phosphatases.
The enzyme's tertiary fold is highly similar to that of
tyrosine-specific phosphatases, except for a 'recognition'
region"
/db_xref="CDD:pfam00782"

ORIGIN

Alignment Scores:

Pred. No.: 3..69e-09 Length: 2264
Score: 20.00 Matches: 20
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.01% Indels: 0
DB: 9 Gaps: 0

US-10-029-345A-109 (1-665) x BC045110 (1-2264)

Qy 242 VALHSCYLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIleMetLys 261
|||||
Db 731 GTCCACTGCTGCTGGCATCTCCGCTCTGCACCATGCCATGCCATCATGAG 790

RESULT 47 AR255944 2377 bp DNA linear PAT 20-DEC-2002
LOCUS AR255944 Sequence 3 from patent US 6482644.
DEFINITION AR255944
ACCESSION AR255944 GI:27305203
KEYWORDS AR255944.1 GI:27305203
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2377)
AUTHORS Cowett, L.M.
TITLE Antisense modulation of dual specific phosphatase 8 expression
JOURNAL Patent: US 6482644-A 3 19-NOV-2002;
FEATURES Location/Qualifiers
source 1..2377
/organism="Unknown"
/mol_type="genomic DNA"

ORIGIN

Alignment Scores:
Pred. No.: 3.84e-09 Length: 2377
Score: 20.00 Matches: 20
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.01% Indels: 0
DB: 6 Gaps: 0

US-10-029-345A-109 (1-665) x AR255944 (1-2377)

Qy 242 VALHSCYLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIleMetLys 261
|||||
Db 864 GTCCACTGCTGCTGGCATCTCCGCTCTGCACCATGCCATGCCATCATGAG 923

RESULT 48 HSU27193 2377 bp mRNA linear PRI 08-DEC-1995
LOCUS HSU27193 Human protein-tyrosine phosphatase mRNA, complete cds.
DEFINITION U27193
ACCESSION U27193
VERSION U27193.1 GI:1109781
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 2377)
AUTHORS Martelli, K.J., Seasholtz, A.F., Kwak, S.P., Clemens, K.K. and
Dixon, J.E.
TITLE hvh-5: a protein tyrosine phosphatase abundant in brain that
inactivates mitogen-activated protein kinase
JOURNAL J. Neurochem. 65 (4), 1823-1833 (1995)
MEDLINE 9609533
PUBMED 7561881
REFERENCE 2 (bases 1 to 2377)
AUTHORS Martelli, K.J.
TITLE Direct Submision
JOURNAL Submitted (16-MAY-1995) Karen J. Martelli, Biochemistry, University
of Michigan, M5416 Medical Science I, 1301 Catherine Street, Ann
Arbor, MI 48109-0606, USA

FEATURES

source

1..2377
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="hvh-5"
/issue_type="brain"
/dev_stage="fetal"
135..2012
/codon_start=1

CDS

/product="protein-tyrosine phosphatase"
/protein_id="AA83151.1"
/db_xref="GI:1109782"
/translation="MAGDRLPRKMDAKKLAISLRGPGPIVDSRSFVEXNSHVL
SSVNTCCSLVKVRLLQKQVTTIELIOPARSOVERTEPDVVVDQSTRDASVLAAD
SELSILSLKLDGCPDSVALITGFATFSSCFPGLCGKPAALLPMSLSPCLPVPVSG
LRLRLPHLYIGSGKVDLDMTONGISVYLNASNSCPDFICSRFMRVPINNYC
EKLLPMLDKSIEFIDKAKLSQCVIYHCLAGISRSATTLAIAYIMKTGMSDDAYRFV
KDRRPSISNPNFNLGOLLVEYRTIKLLALLODPCPTSGTPPPSPAPAGAPLPLRP
PNSRNATNGMAAREBGLSAGSGPPRPFPPTATSAQQCLRHLSRDLQDTNLRK
STSLDTKSAVAPSRKPDGPPDPGAPGLKLDSPSGALDLSSPSPAPAEAR
PPRRRRPPPPASSPAPSPHSLGLNFGDAKQTPRRGLSALAPGLPGPGQAPAGAW
APPLDSPGTSPSPDGWCFSPBGAQAGVLPAPFGAGAPGPGGSDLLRRRRAAARP
RDARTGMPEPAPETQFKRSCOMEPEEGWEGRARGELALGKQASFSGSVEVTEV
S"

ORIGIN

Alignment Scores:
Pred. No.: 3.84e-09 Length: 2377
Score: 20.00 Matches: 20
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.01% Indels: 0
DB: 9 Gaps: 0

US-10-029-345A-109 (1-665) x HSU27193 (1-2377)

QY 242 ValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIleMetLys 261
DB 864 GTTCACTGTGCTGGCATCTCTCGCTGCGCACCATCGCCATCGCATCATGAG 923

RESULT 49
A59887
LOCUS 2415 bp DNA linear PAT 06-MAR-1998
DEFINITION Sequence 5 from Patent WO9706245.
ACCESSION A59887
VERSION A59887.1 GI:3715074
KEYWORDS
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1
AUTHORS Davies, K.E. and Theodosiou, A.
TITLE MAP KINASE PHOSPHATASE GENE AND USES THEREOF
JOURNAL Patent: WO 9706245-A 5 20-FEB-1997;
MEDICAL RES COUNCIL (GB)
FEATURES
source Location/Qualifiers
1..2415
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"

ORIGIN

Alignment Scores:
Pred. No.: 3.89e-09 Length: 2415
Score: 20.00 Matches: 20
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.01% Indels: 0
DB: 6 Gaps: 0

US-10-029-345A-109 (1-665) x A59887 (1-2415)

QY 242 ValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIleMetLys 261
DB 825 GTTCACTGTGCTGGCATCTCTCGCTGCGCACCATCGCCATCGCATCATGAAA 884

RESULT 50
A59888
LOCUS 2453 bp DNA linear PAT 06-MAR-1998
DEFINITION Sequence 6 from Patent WO9706245.
ACCESSION A59888
VERSION A59888.1 GI:3715075

KEYWORDS
SOURCE unidentified
ORGANISM unidentified
unclassified.

REFERENCE 1
AUTHORS Davies, K.E. and Theodosiou, A.
TITLE MAP KINASE PHOSPHATASE GENE AND USES THEREOF
JOURNAL Patent: WO 9706245-A 6 20-FEB-1997;
MEDICAL RES COUNCIL (GB)

FEATURES
source Location/Qualifiers
1..2453
/organism="unidentified"
/mol_type="unassigned DNA"
/db_xref="taxon:32644"

ORIGIN

Alignment Scores:
Pred. No.: 3.94e-09 Length: 2453
Score: 20.00 Matches: 20
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.01% Indels: 0
DB: 6 Gaps: 0

US-10-029-345A-109 (1-665) x A59888 (1-2453)

QY 242 ValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIleMetLys 261
DB 827 GTTCACTGTGCTGGCATCTCTCGCTGCGCACCATCGCCATCGCATCATGAAA 886

Search completed: June 22, 2004, 09:06:14
Job time : 7621 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 22, 2004, 06:28:07 ; Search time 4394 Seconds
(without alignments)
4519.424 Million cell updates/sec

Title: US-10-029-345A-109
Perfect score: 665
Sequence: 1 MAHEMIGTQIVTERLVALLE.....LGVKVSQSSFSGSMRIIEVS 665

Scoring table: OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Word size: 1

Total number of hits satisfying chosen parameters: 200554

Minimum DB seq length: 1419
Maximum DB seq length: 200000000

Post-processing: Listing first 75 summaries

Command line parameters:
-MODE=frame+ p2n.model -DEV=xlh
-Q/cgnt_1/USPTO.spool/US10029345/runat_21062004.164401.29765/app.query.fasta_1.839
-DB=EST -QFMT=fastap -SUFFIX=oligo.rst -MINMATCH=0.1 -LOOPCL=0 -LOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdt -LIST=75
-DOCALLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=50 -MODE=LOCAL -OUTFMT=ptc
-NORM=ext -HEAPSIZ=500 -MINLEN=1419 -MAXLEN=200000000
-USER=US10029345_@CGN_1_1.2810_@runat_21062004.164401.29765 -NCPU=6 -ICPU=3
-NO_MMAB -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :
EST: *
1: em_estba: *
2: em_esthum: *
3: em_estin: *
4: em_estnu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_hic: *
9: gb_est1: *
10: gb_est2: *
11: gb_hic: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estcom: *
17: em_g88_hum: *
18: em_g88_hiv: *
19: em_g88_pln: *
20: em_g88_vit: *
21: em_g88_fun: *
22: em_g88_mam: *
23: em_g88_mus: *
24: em_g88_pro: *
25: em_g88_rtd: *
26: em_g88_phg: *
27: em_g88_vrl: *
28: gb_g88l: *

29: gb_g882: *
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	* Match Length	DB ID	Description
1	67	10.1	3325 11	AK035652
2	42	6.3	2573 11	AK014441
3	20	3.0	2207 11	BC038231
4	12	1.8	2237 11	AK088665
5	12	1.8	2719 11	AK009131
6	12	1.8	2793 11	AK005062
7	12	1.8	2794 11	AK088468
8	9	1.4	1614 11	AK088059
9	9	1.4	1699 11	AK053746
10	9	1.4	2287 11	AK044726
11	9	1.4	2428 11	AK080964
12	8	1.2	1439 28	CC241307
13	8	1.2	1456 13	BU199156
14	8	1.2	1459 29	AG146840
15	8	1.2	1565 11	CNSLTI180
16	8	1.2	1502 12	BG121676
17	8	1.2	1603 28	BZ578841
18	8	1.2	1626 29	AY401953
19	8	1.2	1665 10	BF680537
20	8	1.2	1684 11	AY110929
21	8	1.2	1697 11	AK081315
22	8	1.2	1709 10	BE729007
23	8	1.2	1730 28	BZ576364
24	8	1.2	1742 12	BG859060
25	8	1.2	1746 29	AY400863
26	8	1.2	1770 12	BM320908
27	8	1.2	1792 12	BM321015
28	8	1.2	1839 11	AK045836
29	8	1.2	1867 11	AK017699
30	8	1.2	1918 11	AK075873
31	8	1.2	1919 29	CG755756
32	8	1.2	1965 11	AK047699
33	8	1.2	2008 28	BZ564814
34	8	1.2	2174 10	BF572336
35	8	1.2	2208 10	AW730576
36	8	1.2	2302 11	AK011786
37	8	1.2	2472 11	AY109354
38	8	1.2	2635 11	AK084572
39	8	1.2	2739 11	AK031373
40	8	1.2	3012 11	AK078652
41	8	1.2	3259 11	AK030955
42	8	1.2	3335 11	AK031191
43	8	1.2	3610 11	AK029006
44	8	1.2	3617 11	AK046988
45	8	1.2	3626 11	AK032373
46	8	1.2	3701 11	AK083783
47	8	1.2	3722 11	AK084916
48	8	1.2	3722 11	BC049163
49	8	1.2	3727 11	AK035810
50	8	1.2	3789 11	AK078517
51	8	1.2	4382 11	AK082310
52	8	1.2	4864 11	BC046246
53	8	1.2	1419 11	AK009999
54	7	1.1	1422 12	BG751601
55	7	1.1	1423 28	CC192711
56	7	1.1	1424 29	AG033031
57	7	1.1	1424 29	AY406613
58	7	1.1	1425 11	AK005246
59	7	1.1	1425 13	BQ214023
60	7	1.1	1425 29	AY415850
61	7	1.1	1426 28	CC189843
62	7	1.1	1426 28	CC231606

C 63	7	1.1	1427	13	B0058685
C 64	7	1.1	1428	28	BZ553374
C 65	7	1.1	1429	28	CC241312
C 66	7	1.1	1429	28	CC246343
C 67	7	1.1	1430	28	CC187178
C 68	7	1.1	1433	11	AK053824
C 69	7	1.1	1436	12	BM560037
C 70	7	1.1	1437	29	AY406616
C 71	7	1.1	1438	11	AK040516
C 72	7	1.1	1440	29	CG754464
C 73	7	1.1	1442	11	BC038834
C 74	7	1.1	1442	11	BC048999
C 75	7	1.1	1443	10	BF686707

ALIGNMENTS

RESULT 1
AK035652 3325 bp mRNA linear HNC 19-SRP-2003
LOCUS
DEFINITION Mus musculus adult male urinary bladder cDNA, RIKEN full-length
PHOSPHATASE 7 homolog [Homo sapiens], full insert sequence.
ACCESSION AK035652.1 GI:26330815
VERSION AK035652
KEYWORDS HNC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE

1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
JOURNAL MEDLINE 99279253
PUBMED 10349636
REFERENCE AUTHORS

TITLE

Itch, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
JOURNAL MEDLINE 20499374
PUBMED 11042159
REFERENCE AUTHORS

TITLE

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, B., Watabiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)
JOURNAL MEDLINE 20530913
PUBMED 11076861
REFERENCE AUTHORS

TITLE

The RIKEN Genome Exploration Research Group Phase II Team and the
PANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)
JOURNAL MEDLINE 11076861
PUBMED 11076861
REFERENCE AUTHORS

TITLE

The PANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
JOURNAL MEDLINE 11076861
PUBMED 11076861
REFERENCE AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hitozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,

Karich, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koye, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Nunakazi, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akashira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toy, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.

COMMENT

Direct Submission
Submitted (16-JUN-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.

FEATURES

source

1..3325
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM DB:9530081F05"
/db_xref="MG1:2399973"
/db_xref="taxon:10090"
/clone="9530081F05"
/sex="male"
/tissue_type="urinary bladder"
/clone_id="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
460..1080
/note="unnamed protein product, TRUNCATED MAPK PHOSPHATASE
7 homolog [Homo sapiens] (SPTK19) [Q96Q52, evidence: FASTA,
96.7%id, 86%length, match=369]
putative"
/codon_start=1
/protein_id="BAC29138.1"
/db_xref="GI:26330816"
/translation="MAHEMTGTQIVTSVALVLSGTEKVLIDSRPEVYNTSHILE
AINTNCSLMRRLQODKVLITELIHSARKVIDCNQVVDQSDQVDSLSDC
FLVTILGLTSLSPVSHLILAGFAEPSPFCPLCEKSTLPTCISQCLPVAINGPT
RIPLNVLGCRDVLNVKVASKHLISLCSVASGRAAVVHYLSMIL"

CDS

ORIGIN

Alignment Scores:

Pred. No.: 1.86e-52 Length: 3325
Score: 67.00 Matches: 160
Percent Similarity: 96.39% Conservative: 0
Best Local Similarity: 96.39% Mismatches: 3
Query Match: 10.08% Indels: 6
DB: 11 Gaps: 0

US-10-029-345a-109 (1-665) x AK035652 (1-3325)

QY 15 leuValAlaIeuIeuGluSerGlyThrGluIuYValIeuIeuIleAspSerArgProPhe 34
DB 502 TTGGTGCTCTGCTGGAAAGTGAACGAAAGAGTGGTAATGTAAGCGACCAATT 561
QY 35 ValGluTYraAntTrSerHisIleIeuGluIuAlaIleAsnIleAsnCySerIysIeuMet 54
DB 562 GTGAATKATCATAGCTCCACATTTTGGAGCCATTAAATCAACTGCTCCAACTGATG 621
QY 55 LysArgArgIeuGluIngIuAspIysValIeuIleThrGluIeuIleGlnHisSerAlaIys 74
DB 622 AAGCGAAGTGGCAACGAAAGTAATTAATTACAGAACTTATTCACCAATTCGCAAG 681

QY 75 HsLysValAspIleAspCySer-GlnHisValValValTyAspGlnSerSerGlnAs 94
Db 682 CAAAGAGTTGACATTCGTCGCA-TCGAAGAGGTGATGATTATGATCAAAAGTCCCAAG 740
QY 94 PValAla-SerLeuSerSerAspCyPheLeuThrValLeuLeuGlyLysLeuGlnLys- 113
Db 741 TGTGTG-TTCTCTGTGTGTCAGACTGCTTCTCTCACTGACTCTCGGGAACGAGAG-G 798
QY 114 SerPheAsnSerValHisLeuLeuAlaGlyGlyPheAlaGlnPheSerArgCysPhePro 133
Db 799 AGTTCAACTCTCTCCACTGCTGAGGGGCTTGCTGAGTTCCTGCTGTTCTCCCT 858
QY 134 GlyLeuCyGlnGlyLysSerThrLeuValProThrCysIleSerGlnProCysLeuPro 153
Db 859 GGCCTCTGGAAGAAAGTCATCTAGTCCCACTGATCTGATCTGAGCCTTCTTACT 918
QY 154 ValAlaAsnIleGlyProThrArgIleLeuProAsnLeuTyLeuGlyCysGlnArgAsp 173
Db 919 GTTGCAACACTTGCGCACTCGAATCTTCCCAATCTATCTTGCTGCGCAGCAGAT 978
QY 174 ValLeuAsnLys 177
Db 979 GTCTCAACAG 990
RESULT 2
AK014441 2573 bp mRNA linear HTC 20-SEP-2003
LOCUS AK014441
DEFINITION Mus musculus 18 days pregnant adult female placenta and extra
embryonic tissue cDNA, RIKEN full-length enriched library,
clone:3830417M17 product:RUNCATED MAPK PHOSPHATASE 7 homolog [Homo
sapiens], full insert sequence.
ACCESSION AK014441
VERSION AK014441.1 GI:12852294
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Carninci,P. and Hayashizaki,Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE
AUTHORS 3
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
Kono,H., Akiyama,J., Nishi,K., Katsunai,T., Taahito,H., Itoh,M.,
Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
Yamamoto,K., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Wataniki,M.,
Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE
AUTHORS 4
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
REFERENCE
AUTHORS 5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.

TITLE Analysis of the mouse transcriptome based on functional annotation
JOURNAL of 60,770 full-length cDNAs
REFERENCE Nature 420, 563-573 (2002)
AUTHORS 5 (bases 1 to 2573)
Abechi,J., Aizawa,K., Akahira,S., Akimura,T., Arai,A., Aono,H.,
Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Fukunishi,Y.,
Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hiramoto,K.,
Hirooka,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Izawa,M.,
Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Kono,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K.,
Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C.,
Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T., Sogabe,Y.,
Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T.,
Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A., Yoshida,K.,
Yoshino,M., Yamatsu,M. and Hayashizaki,Y.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suenho-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
URL: <http://genome.gsc.riken.go.jp/>, Tel:81-45-503-9222,
Fax:81-45-503-9216)
COMMENT Please visit our web site (<http://genome.gsc.riken.go.jp/>) for
further details.
CDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5' GAGGAGAGAGAGATCCAGAGCTTTTCTTTTCTTTTNN 3']. cDNA was
prepared by using trehalose thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. cDNA went
through one round of normalization to Rot = 10.0 and subtraction to
Rot = 50.0. Second strand cDNA was prepared with the primer adapter
of sequence [5'
GAGGAGAGATTCGAGTTATTAATTAATCCCCCCCCC 3']. cDNA was cleaved
with BamHI and XhoI. Vector: a modified pBlueScript KS(+) after
bulk excision from Lambda FLC 1. Cloning sites, 5' end: SalI; 3'
end: BamHI. Host: DH10B.
Location/Qualifiers
FEATURES
source
1..2573
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:3830417M17"
/db_xref="MGI:1903116"
/db_xref="taxon:10090"
/clone="3830417M17"
/sex="female"
/tissue_type="placenta and extra embryonic tissue"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="18 days pregnant adult"
1..2573
/note="TRUNCATED MAPK PHOSPHATASE 7 homolog [Homo sapiens]
[SPR]Q96Q52, evidence: FASTA, 96.7%id, 86%length,
match=369]"
ORIGIN
Alignment Scores:
pred. No.: 1 22e-28 Length: 2573
Score: 42.00 Matches: 42
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 6.32% Indels: 0
DB: 11 Gaps: 0
US-10-029-345A-109 (1-665) x AK014441 (1-2573)
QY 624 TyAspArgSerCyGlnMetGlnPheGlyGlnSerIleMetSerGlnAsnArgSerArg 643
Db 41 AAACGAGAGAGCTGCCAAATTCGATTTGAGAGAGAGATTTATGTCGAGAAACGCTCCAG 100

OY	664	Gluc1IuEnGjLYxvVa1Gj5eCtIuSeRSeRPhSeRgJysenMeCtIu1Ic1gJ1	663
DB	101	GAGGAGCTGGGCGAAGCTGGGCGACCGACCTTCGCGGAGCATGAGATCATTCGAG	160
OY	664	ValSer 665	
DB	161	GTCTCT 166	
RESULT 3			
LOCUS	BC038231	2207 bp	mRNA
DEFINITION	Human sapiens, similar to dual specificity phosphatase 8, clone		
ACCESSION	BC038231		
VERSION	BC038231.1		
KEYWORDS	HTC.		
SOURCE	Human sapiens (human)		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 2207)		
AUTHORS	Straussberg, R.		
TITLE	Submitted (30-SEP-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
JOURNAL	NIH-MGC Project URL: http://mgc.nci.nih.gov		
REMARK	Contact: MGC help desk		
COMMENT	Email: cgaphs-remail.nih.gov Tissue Procurement: ATCC/DCTD/DTF cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland; Web site: http://www.nisc.nih.gov/ Contact: nisc.mgc@nih.gov		
	Akhter, N., Ayale, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Grant, S., Guan, X., Gupta, V., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Latic, P., Legaspi, R., Maduro, Q.N., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stancirip, S., Thomas, P.J., Touchman, J.W., Tsurguen, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, H., and Green, E.D.		
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IMAX Plate: 79 Row: a Column: 12 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA g1: 4758211 This clone has the following problem: frame shifted.		
FEATURES	Location/Qualifiers		
source	1..2207		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:5547764"		
	/tissue_type="Skin, melanotic melanoma."		
	/clone_lib="NIH MGC_72"		
	/lab_host="DH10B"		
	/note="Vector: pCMV-SPORT6"		
ORIGIN			
Alignment Scores:			
Pred. No.:	1,19e-07	Length:	2207
Score:	20.00	Matches:	20
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.01%	Indels:	0
DB:	11	Gaps:	0

UTS-10-029-345A-109 (11-665) X BC038231 (11-2207)	
By	242 ValhiCysLeuAlaGlyTLeSerArgSerAlaThrIleAlaIleAlaTyrIleMetIys 261
Db	861 GTCACCTGCTCTGGCTGGCATCTCCGCTCTGCACACATCGCATCGGCTCATCATGTAAG 920
RESULT 4	
AK088665	2237 bp mRNA linear HTC 20-SEP-2000
LOCUS	
DEFINITION	Mus musculus 2 days neonate thymus thymic cells cDNA, RIKEN
ACCESSION	specificity enriched library, clone: E430023B14 product: dual
VERSION	specificity phosphatase 6, full insert sequence.
KEYWORDS	AK088665
SOURCE	AK088665.1 GI:26353717
ORGANISM	HTC; CAP trapper.
	Mus musculus (house mouse)
	Mus musculus
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;
REFERENCE	
AUTHORS	1
TITLE	Carninci, P. and Hayashizaki, Y.
JOURNAL	High-efficiency full-length cDNA cloning
MEDLINE	Meth. Enzymol. 303, 19-44 (1999)
PUBMED	99279253
REFERENCE	10349636
AUTHORS	2
TITLE	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
JOURNAL	Itoh, M., Komoto, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
MEDLINE	Normalization and subtraction of cap-trapper-selected cDNAs to
PUBMED	prepare full-length cDNA libraries for rapid discovery of new gene
REFERENCE	Genome Res. 10 (10), 1617-1630 (2000)
AUTHORS	20499374
TITLE	11042159
JOURNAL	
MEDLINE	
PUBMED	
REFERENCE	
AUTHORS	3
TITLE	Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P.,
JOURNAL	Komoto, H., Akiyama, J., Nishi, K., Kitenunai, T., Tashiro, H., Itoh, M.,
MEDLINE	Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
PUBMED	Yamanoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
REFERENCE	Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubiki, M.,
AUTHORS	Yoneda, Y., Ishikawa, K., Ozawa, K., Tanaka, T., Matsumura, S., Kawai, J.
JOURNAL	Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
MEDLINE	RIKEN integrated sequence analysis (RISA) system--384-format
PUBMED	sequencing pipeline with 384 multipipillary sequencer
REFERENCE	Genome Res. 10 (11), 1757-1771 (2000)
AUTHORS	20530913
TITLE	11076861
JOURNAL	
MEDLINE	
PUBMED	
REFERENCE	
AUTHORS	4
TITLE	THE RIKEN Genome Exploration Research Group Phase II Team and the
JOURNAL	PANTOM Consortium.
MEDLINE	Functional annotation of a full-length mouse cDNA collection
PUBMED	Nature 409, 695-690 (2001)
REFERENCE	
AUTHORS	5
TITLE	The PANTOM Consortium and the RIKEN Genome Exploration Research
JOURNAL	Group Phase I & II Team.
MEDLINE	Analysis of the mouse transcriptome based on functional annotation
PUBMED	of 60,770 full-length cDNAs
REFERENCE	Nature 420, 563-573 (2002)
AUTHORS	6 (bases 1 to 2237)
TITLE	Adachi, U., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
JOURNAL	Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
MEDLINE	Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T.,
PUBMED	Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
REFERENCE	Kato, H., Kawai, J., Kojima, Y., Kondo, S., Komoto, H., Kouda, M.,
AUTHORS	Koyama, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
JOURNAL	Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Onno, M., Ohnato, N.,
MEDLINE	Okazaki, Y., Saito, R., Saitoh, K., Sakai, C., Sakai, K., Sakazume, N.,
PUBMED	Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T.,
REFERENCE	Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
AUTHORS	Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
JOURNAL	Muramatsu, M. and Hayashizaki, Y.
MEDLINE	Direct Submision
PUBMED	Submitted (16-Apr-2002) Yoshinide Hayashizaki, The Institute of

with XhoI and SctI. Cloning sites, 5' end: XhoI; 3' end: SctI.
Host: SOUR.

FEATURES
source
Location/Qualifiers

1..2719
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:2310004A03"
/db_xref="MGI:1899068"
/db_xref="taxon:10090"
/clone="2310004A03"
/sex="male"
/tissue_type="tongue"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
390..1535
/note="unnamed protein product; dual specificity phosphatase 6 (MGD|MGI:1914853, GB|NW_026268, evidence: BLASTN, 9%, match=2793)
putative"

CDS

/codon_start=1
/protein_id="BAB26093.1"
/db_xref="GI:12843735"
/translation="MIDTLRPVPPFASMAICTVSGLANOLEIENGEPFLMDCRPOEL
YESSHIESAINVAIPGIMLRLOKGNLPYRALFTCEDEDFRRCGTDTVVLVDENS
SPWNETGSESVGLILKLIKDEGCFALFEGFSKFOEPALHGEITLNDGSCSSSP
PVLVLGIRISDSSDIESLDLPDSDPSLSDGSPSPVPEIILPEVLYGC
ADSTNLDTLVEERGICILANTVNTPLPENNAGEPKYKQIPISDMISQSLSPFEPAI
SPIDARKGRCCVYHCLAGISRSVTVTAYAIMQKLNLSWNAVYIVKKKSNISPNF
NFMGLDPERLTGLSSPCDNRPVLPOLFTTSPNONVYQVDSLOST"
2700..2705
/note="putative"
2719
/note="putative"
/note="putative"

ORIGIN

Alignment Scores:
Pred. No.: 6.78 Length: 2719
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservatave: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: 11 Gaps: 0

US-10-029-345A-109 (1-665) x AK009131 (1-2719)

QY 240 ValLeuValHisCysLeuValaGlyTylLeSerArgSer 251

Db 1254 GTCTGTGTCATTGCTGTGACGATACAGCCCTCT 1289

RESULT 6
AK005062 2793 bp mRNA linear HTC 20-SEP-2003
LOCUS
DEFINITION Mus musculus adult male liver cDNA, RIKEN full-length enriched
library, clone:1300019103 product:dual specificity phosphatase 6,
full insert sequence.

ACCESSION AK005062
VERSION AK005062.1 GI:12836727
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
1 Carninci, P. and Hayashizaki, Y.
Autors TITLE High-efficiency full-length cDNA cloning
JOURNAL Mech. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Kono, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kaishiwagi, K.,
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multipipillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

TITLE

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

TITLE
JOURNAL
REFERENCE
AUTHORS
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2793)

JOURNAL
REFERENCE
AUTHORS
Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H.,
Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,
Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Himoto, K.,
Hiraoka, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M.,
Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K.,
Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C.,
Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D.,
Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y.,
Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T.,
Tejima, Y., Toya, T., Yamamura, T., Yasunishi, A., Yoshida, K.,
Yoshino, M., Muramatsu, M. and Hayashizaki, Y.
Direct Submission

TITLE
JOURNAL
Submitted (10-JUL-2000) Yoshihide Hayashizaki. The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

COMMENT
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.

cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5' GAGAGAGAGACGCGCCGCACTCCAGTCTTTTCTTTTCTTAA 3'], cDNA was
prepared by using trenaiose thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. Second
strand cDNA was prepared with the primer adapter of sequence [5'
GAGAGAGAGAGATCCCAAGAGCTCAATTAATTAATTAACCCGCCCC 3']. cDNA was
cleaved with XhoI and SctI. Cloning sites, 5' end: SctI; 3' end:
XhoI. Host: SOUR.

FEATURES
Location/Qualifiers

1..2793
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:1300019103"
/db_xref="MGI:1897200"
/db_xref="taxon:10090"
/clone="1300019103"
/sex="male"

PLPLVIGLGLIRISSSSDISSLDLDRDPNSATDSGLSPNSOPFPVELIPLVILG
 AKDSTNLDVLEEFGIKYIANTPNL.PMLFENAGEFKXKO.PISPHWSON.SOPFPEAI
 SFIDEARGKNGV.VHCLAGSR.SVTTVAYVKNKALNSMADYDKAKSKNSIPWF
 NEMQQLDFERTLCLSSPCDNKRVTPOLYFTTBSNOWNVYDLSQST"
 2775..2780
 /note="putative"
 2794
 /note="putative"
 2794
 /note="putative"
 ORIGIN
 Alignment Scores:
 Pred. No.: 6.98 Length: 2794
 Score: 12.00 Matches: 12
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.80% Indels: 0
 DB: 11 Gaps: 0
 US-10-029-345A-109 (1-665) x AK088468 (1-2794)
 Oy 240 ValIeuValH3CysLeuAlaGlyIleSerArgSer 251
 |||||
 Db 1327 GTCTGTGATGCTCTGTGGCAGGATACGCCGCTCT 1362
 RESULT 8 1614 bp mRNA linear HTC 20-SEP-2003
 AK088059
 LOCUS Mus musculus 2 days neonate thymus thymic cells cDNA, RIKEN
 DEFINITION full-length enriched library, clone:E430002P08 product:dual
 specificity phosphatase 2, full insert sequence.
 ACCESSION AK088059.1 GI:26104661
 VERSION AK088059.1
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 REFERENCE
 1 Carninci,P. and Hayashizaki,Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636
 2 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
 Itoh,M., Kohno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
 JOURNAL prepare full-length cDNA libraries for rapid discovery of new genes
 MEDLINE 20499374
 PUBMED 11042159
 3 Shibata,K., Itoh,M., Aizawa,K., Nagaoaka,S., Sasaki,N., Carninci,P.,
 Kohno,H., Akiyama,J., Nishi,K., Kitsuana,T., Tashiro,H., Itoh,M.,
 Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
 Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
 Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,B., Watabiki,M.,
 Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
 Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
 TITLE RIKEN integrated sequence analysis (RISA) system-384-format
 JOURNAL sequencing pipeline with 384 multichipillary sequencer
 MEDLINE 20530913
 PUBMED 11076861
 4 The RIKEN Genome Exploration Research Group Phase II Team and the
 JOURNAL FANTOM Consortium.
 TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL Nature 409, 685-690 (2001)
 REFERENCE 5
 AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research
 TITLE Group Phase I & II Team.
 ANALYSIS Analysis of the mouse transcriptome based on functional annotation

JOURNAL of 60,770 full-length cDNAs
 REFERENCE Nature 420, 563-573 (2002)
 6 (bases 1 to 1614)
 AUTHORS Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
 Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
 Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraka,T., Hirozane,T.,
 Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kaikawa,T.,
 Kato,H., Kawai,D., Kojima,Y., Kondo,S., Kohno,H., Konda,M.,
 Koyama,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
 Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
 Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
 Sano,H., Sasaki,D., Shibata,K., Shingawa,A., Shiraki,T.,
 Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akhiba,S.,
 Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
 Muramatsu,M. and Hayashizaki,Y.
 TITLE Direct Submission
 JOURNAL Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute; 1-7-22 Suenho-cho, Tsukuba-shi, Yokohama,
 Kanagawa 230-0045, Japan [E-mail:genome-res@gscc.riken.go.jp,
 URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
 Fax:81-45-503-9216]
 COMMENT cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Tissues were provided by Dr. John Todd (Dept. of Medical Genetics
 Wellcome Trust Centre for Molecular Mechanisms in Disease Wellcome
 Trust/MRC building Addenbrookes Hospital Cambridge) whose
 assistance we gratefully acknowledge.
 Please visit our web site for further details.
 URL:http://genome.gsc.riken.go.jp/
 URL:http://fantom.gsc.riken.go.jp/
 location/Qualifiers
 1..1614
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="NOD"
 /db_xref="FANTOM DB:E430002P08"
 /db_xref="MG1:2426956"
 /db_xref="taxon:10090"
 /clone="E430002P08"
 /cell_type="thymic cells"
 /tissue_type="thymus"
 /clone_id="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="2 days neonate"
 93..1032
 /note="dual specificity phosphatase 2 (GB|U11330,
 evidence:BLASTN, 98%, match=1511)
 putative"
 ORIGIN
 Alignment Scores:
 Pred. No.: 2.9e+03 Length: 1614
 Score: 9.00 Matches: 9
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.35% Indels: 0
 DB: 11 Gaps: 0
 US-10-029-345A-109 (1-665) x AK088059 (1-1614)
 Oy 246 AlaGlyIleSerArgSerAlaThrIle 254
 |||||
 Db 862 GGTGGCATCTCTCGCTCAGCCACCATC 888
 RESULT 9 1699 bp mRNA linear HTC 20-SEP-2003
 AK053746
 LOCUS Mus musculus 0 day neonate eyeball cDNA, RIKEN full-length enriched
 DEFINITION library, clone:E130306H24 product:DUAL SPECIFICITY PROTEIN

PHOSPHATASE 4 (EC 3.1.3.48) (EC 3.1.3.16) (MITOGEN-ACTIVATED PROTEIN KINASE PHOSPHATASE-2) (MAP KINASE PHOSPHATASE-2) (MKP-2) homolog [Rattus norvegicus], full insert sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AK053746.1 GI:26343694
AK053746.1 GI:26343694
HTC; CAP trapper.
Mus musculus (house mouse)
Mus musculus

REFERENCE

1 Carninci, P. and Hayashizaki, Y.
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

TITLE

High-efficiency full-length cDNA cloning

JOURNAL

Meth. Enzymol. 303, 19-44 (1999)

PUBMED

99279253
10349636

REFERENCE

2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL

Genome Res. 10 (10), 1617-1630 (2000)

PUBMED

20493374
11042159

REFERENCE

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishii, K., Kiteunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujikawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsumura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384 format sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL

Genome Res. 10 (11), 1757-1771 (2000)

PUBMED

20530913
11076661

TITLE

The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

JOURNAL

Nature 409, 685-690 (2001)

PUBMED

20530913
11076661

REFERENCE

4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

TITLE

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

JOURNAL

Nature 420, 563-573 (2002)

PUBMED

20530913
11076661

REFERENCE

5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

TITLE

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hamagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, N., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishii, K., Nomura, K., Numasaki, K., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Kahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.

JOURNAL

Nature 420, 563-573 (2002)

PUBMED

20530913
11076661

REFERENCE

6 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

TITLE

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hamagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, N., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishii, K., Nomura, K., Numasaki, K., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Kahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.

JOURNAL

Nature 420, 563-573 (2002)

PUBMED

20530913
11076661

REFERENCE

7 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

TITLE

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hamagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, N., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishii, K., Nomura, K., Numasaki, K., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Kahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.

JOURNAL

Nature 420, 563-573 (2002)

PUBMED

20530913
11076661

REFERENCE

8 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

Prepare mouse tissues.
Please visit our web site for further details.
URL: <http://genome.gsc.riken.go.jp/>
URL: <http://fantom.gsc.riken.go.jp/>

FEATURES

source

1.1699
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM:DB:E130306H24"
/db_xref="MGI:2425595"
/db_xref="taxon:10090"
/clone="E130306H24"
/issue_type="eyeball"
/clone_id="RIKEN full-length enriched mouse cDNA library"
/dev_stage="0 day neonate"
117.1313

CDS

/note="unnamed protein product; DUAL SPECIFICITY PROTEIN PHOSPHATASE 4 (EC 3.1.3.48) (EC 3.1.3.16) (MITOGEN-ACTIVATED PROTEIN KINASE PHOSPHATASE-2) (MAP KINASE PHOSPHATASE-2) (MKP-2) homolog [Rattus norvegicus] (SWISSPROT|062767, evidence: FASTA, 98.5%id, 100%length, match=1194)
putative"
/codon_start=1
/protein_id="BAC35504.1"
/db_xref="GI:26343695"
/translation="MTWTELEEMDCSVLRKMRNDENGCGGAGNGSGHGLGLSGCKLLDCRPFLLHSAAGVIRGSVVRNCNTIVRRKAKSVSLQILPAEBEVRARL SGVSAVIVYDRSPRAESLRSDSTVSLVQALRRARERDTICLGGYERFSEYPS FSKTKALAIIPVPVPTNESLIDGSCSGPVLDOGGPVELPLVGSAYHVARC DMVLDIGTALAINVSDCPNHFEGYQYCIYEDNHRKADISSMFEALREYDAVDC RRRVLVHCQAGSRSLTITLALMKKRVRLBAEFVKORSIISPNPSCQLQF ESQVLTSCAABAASPSGPLRRERKATPTPTSPFVSFVSQVMAAPNLPLHSPITLSPSC"

ORIGIN

Alignment Scores:

Pred. No.: 3.06e+03 Length: 1699
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.35% Indels: 0
DB: 11 Gaps: 0

US-10-029-345A-109 (1-665) x AK053746 (1-1699)

Oy 246 AAGAGYILESerfgrSeraLaThrile 254

Db 972 GCCGCAATCTCCGATCAGCCACCATC 998

RESULT 10

AK044726

LOCUS

Mus musculus adult retina cDNA, RIKEN full-length enriched library, clone:AB30037D12 product:hypothetical protein, full insert

DEFINITION

AK044726 2297 bp mRNA linear HTC 20-SEP-2003

ACCESSION

AK044726

VERSION

AK044726.1 GI:26336738

KEYWORDS

HTC; CAP trapper.

SOURCE

Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)

AUTHORS

Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL MEDLINE Genome Res. 10 (10), 1617-1630 (2000)

PUBMED 20499374

REFERENCE 11042159

AUTHORS 3
Shibata,K., Itoh,M., Aizawa,K., Nagaoaka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuami,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujitake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.,

TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer

JOURNAL MEDLINE Genome Res. 10 (11), 1757-1771 (2000)

PUBMED 20530913

REFERENCE 11076861

AUTHORS 4
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

TITLE Functional annotation of a full-length mouse cDNA collection

JOURNAL MEDLINE Nature 409, 685-690 (2001)

PUBMED 11706861

REFERENCE 11076861

AUTHORS 5
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL MEDLINE Nature 420, 563-573 (2002)

PUBMED 1202297

REFERENCE 1202297

AUTHORS 6
Fukuchi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Adachi,S., Furuta,M., Hanagaki,T., Hara,A., Hashizume,M., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiroaka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Katsukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Komori,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakatsume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.

TITLE Direct Submission

JOURNAL MEDLINE Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Kanagawa 230-0045, Japan (E-mail:genome-res@gsr.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in RIKEN Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

FEATURES Retina RNA was provided by Dr. Stefano Gustincich (Department of Neurobiology, Harvard Medical School, 220 Longwood Ave., Boston, MA 02115, USA) whose assistance is gratefully acknowledged. Please visit our web site for further details.

URL: http://genome.gsc.riken.go.jp/

location/Qualifiers 1..2297

source /organism="Mus musculus"

mol_type "mRNA"

strain "C57BL/6J"

db_xref="PANTOM,DB:A930037D12"

db_xref="MGI:2410058"

db_xref="taxon:10090"

clone "A930037D12"

tissue_type "retina"

clone_lib="RIKEN full-length enriched mouse cDNA library"

dev_stage="adult"

CDS

<1..571

/note="unnamed protein product; hypothetical protein (evidence: ProCrest,NCBI CDS Predictor) putative"

/codon_start=2

/protein_id="BAC32052.1"

/db_xref="GI:26336739"

/translation="QDSPATLTGRARAAARSRVKEKIQLRKDPSPRAGVGLPPSG SPGRAGCGMWRVGLGRGASGCCGPTTGGVREGLHAYSGAPHPIDFVTPPEP LGTSLERHAADDLQPRRGSQDPRRAAGTIGLRSGILGAPCSQARPPPPAVQA HDCTGGQGPGRDLGPGRAASSLSAUSL"

ORIGIN

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
1	9.00	4.22e+03	9	0	0	0	0
2	100.00%	100.00%	0	0	0	0	0
3	100.00%	100.00%	0	0	0	0	0
4	1.35%	1.35%	0	0	0	0	0

DB: 11

US-10-029-345A-109 (1-665) x AK044726 (1-2297)

Qy 495 AlaGlnArgSerLeuLeuSerProLeu 503

Db 1624 GCACAGAGAGAGCGCTTTTGAGTCCTTTA 1650

RESULT 11

AK080964 2428 bp mRNA linear HTC 20-SEP-2003

LOCUS Mus musculus 4 days neonate male adipose cDNA, RIKEN full-length enriched library, clone:BA30220P3 product:DUPL SPECIFICITY PROTEIN PHOSPHATASE 4 (EC 3.1.3.48) (EC 3.1.3.16) (MITOGEN-ACTIVATED PROTEIN KINASE PHOSPHATASE-2) (MAP KINASE PHOSPHATASE-2) (MKP-2) homolog (Rattus norvegicus), full insert sequence.

ACCESSION AK080964

VERSION AK080964.1 GI:26348914

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS Carninci,P. and Hayashizaki,Y.

TITLE High-efficiency full-length cDNA cloning

JOURNAL MEDLINE Meth. Enzymol. 303, 19-44 (1999)

PUBMED 99279253

REFERENCE 10349636

AUTHORS 2
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL MEDLINE Genome Res. 10 (10), 1617-1630 (2000)

PUBMED 20499374

REFERENCE 11042159

AUTHORS 3
Shibata,K., Itoh,M., Aizawa,K., Nagaoaka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuami,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujitake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.,

TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer

JOURNAL MEDLINE Genome Res. 10 (11), 1757-1771 (2000)

PUBMED 20530913

REFERENCE 11076861

AUTHORS 4
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

TITLE Functional annotation of a full-length mouse cDNA collection

JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Nature 409, 685-690 (2001)

5
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 2428)

Naichi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carrinci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Komori, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Ohnato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.

Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

CNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.
Location/Qualifiers

1. 2428
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM.DB:B430220P13"
/db_xref="MG1:2412162"
/db_xref="taxon:10090"
/clone="B430220P13"
/sex="male"
/tissue_type="adipose"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="4 days neonate"
120. 1316
/note="unnamed protein product; DUAL SPECIFICITY PROTEIN PHOSPHATASE 4 (EC 3.1.3.48) (EC 3.1.3.16)
(MITOGEN-ACTIVATED PROTEIN KINASE PHOSPHATASE-2) (MAP KINASE PHOSPHATASE-2) (MKP-2) homolog [Rattus norvegicus] (SWISSPROT|Q62767, evidence: FASTA, 98.5%id, 100%length, match=1194)
putative"
/codon_start=1
/protein_id="BAC38097.1"
/db_xref="GI:26348915"
/translation="MYMELREMDGCVLRKLRNDRNGGSGNGSGSHALGILSGGKCLLDRCRPLFASAGITRGSVNRCTTVRRRKGVSLEQILPAEEVRARLRSGLSAVIVYDERSRAESLRDSTVSLVQALRRNARDIDILKGGYRFSSEYEFCSKTKALAIIPPPVPSTNEDLDGSGCGTPIHQGQVEIILPVLGSAVHAARDMLDALGITALLNNSDCPNHFEHGYKCIPEVDNKKADISSMFEALPIYIDAVDCRGRLVHCGAGISRSATICTLAVLMKKRVLSEAFEVKORRGIISNPSFMGQLQFESQVLTTSCAAELASPSGRLRERKATPPTSGVFSPVSVGHAAPSLPLHSEITTSRSC"

polya_signal
2404. 2409
/note="putative"
2428
/note="putative"
ORIGIN

Alignment Scores:
Pred. No.: 4,48e+03 Length: 2428
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.35% Indels: 0
DB: 11 Gaps: 0

US-10-029-345a-109 (1-665) x AK080964 (1-2428)

QY 246 AAG1Y115sera7sera1athrile 254
|||||
Db 975 GCCGCATCTCCGCATCAGCCACCATC 1001
|||||

RESULT 12
CC241307 1439 bp DNA linear GSS 12-MAY-2003
CC241307/c
CH261-10J7 Sp6.1 CH261 Gallus gallus genomic clone CH261-10J7,
LOCUS
DEFINITION
Genomic survey sequence.
ACCESSION
CC241307
VERSION
CC241307.1 GI:30567970
KEYWORDS
GSS.
SOURCE
Gallus gallus (chicken)
ORGANISM
Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
1 (bases 1 to 1439)
REFERENCE
AUTHORS
Kremetzki, C., Higginbotham, J., Wylie, K., Carter, J., McPherson, J., Warren, W., Graves, T., Mardis, E. and Wilson, R.
Gallus gallus BAC End Reads
Unpublished (2003)
CONTACT: Richard K. Wilson
Genome Sequencing Center
Washington University School of Medicine
Email: submissions@wustl.edu
Insert Length: 18200 Std Error: 0.00
Seq primer: Sp6 ATTAGTCACTATAG
Class: BAC ends
High quality sequence start: 205
High quality sequence stop: 561.
Location/Qualifiers

1. 1439
/organism="Gallus gallus"
/mol_type="genomic DNA"
/strain="Red Jungle Fowl"
/db_xref="taxon:9031"
/clone="CH261-10J7"
/sex="female"
/cell_line="UCD001, inbred 256"
/clone_lib="CH261"
/note="Vector: pTARBAC2.1; Site 1: EcoRI; Site 2: EcoRI; CH261 Female Chicken library - For library and clone ordering information: http://www.chori.org/bacpac"

ORIGIN

Alignment Scores:
Pred. No.: 2,32e+04 Length: 1439
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 28 Gaps: 0

US-10-029-345a-109 (1-665) x CC241307 (1-1439)

QY 359 Serva1ProSerValProSerVal 366
|||||
Db 243 TCAGTTCCTTCTGTGCTTCAGTT 220
|||||

RESULT 13
BU199156

LOCUS BUI99156 1456 bp mRNA linear EST 05-SBP-2002
 DEFINITION DBCID04 DCB Homo sapiens cDNA, mRNA sequence.
 ACCESSION BUI99156
 VERSION BUI99156.1 GI:22717746
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 1456)
 Xu,X., Gu,J., Liu,F., Qu,J., Zhao,M., Li,Y., Huang,Q., Zhou,J., Song,H., Gu,Y., Yang,Y., Gao,G., Xiao,H., Li,N., Qian,B., Gao,X., Cheng,Z., Xu,S., Gu,W., Tu,Y., Jia,J., Fu,G., Ren,S., Zhong,M., Lu,G., Cheng,Z., and Han,Z.
 Homo sapiens cDNA DCB clones
 Unpublished (2000)
 CONTACT: Zeguang Han
 Chinese National Human Genome Center at Shanghai
 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
 201203, P. R. China
 Tel: 86-21-50801919(ex.45)
 Fax: 86-21-50801922
 Email: hanzg@chgc.sh.cn.
 FEATURES
 source
 1..1456
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /cell_type="dendritic cells"
 /dev_stage="mature"
 /lab_host="BM25.8"
 /clone_lib="DCB"
 /note="Vector: pTritpLx2; Site_1: sfilA; Site_2: sfilB"
 ORIGIN
 Alignment Scores:
 Pred. No.: 2.35e+04 Length: 1456
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.20% Indels: 0
 DB: 13 Gaps: 0
 US-10-029-345A-109 (1-665) x BUI99156 (1-1456)
 QY 363 ValProSerValGlnProSerLeu 370
 |||||
 Db 355 GTCCCATCTGTCACACCTCTCTTG 378
 RESULT 14
 AG146840 1459 bp DNA linear GSS 08-JUN-2002
 LOCUS AG146840
 DEFINITION Pan troglodytes DNA, clone: RP43-008124.TU, genomic survey sequence.
 ACCESSION AG146840
 VERSION AG146840.1 GI:16676518
 KEYWORDS GSS.
 SOURCE Pan troglodytes (chimpanzee)
 ORGANISM Pan troglodytes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Pan.
 1
 Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Tockki,Y., Watanabe,H. and Sakaki,Y.
 BAC end sequences of Library RP43-008124
 Unpublished
 2 (bases 1 to 1459)
 Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T., Tockki,Y., Watanabe,H. and Sakaki,Y.
 Direct Submission
 Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center (GSC); 1-7-22 Suenhiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

(E-mail: chimpanzee@sc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/, Tel: 81-45-503-9111, Fax: 81-45-503-9170)
 Clones are derived from the chimpanzee BAC library RP43-008124. This BAC end was generated during the R&D process and may have higher chance of clone tracking errors.
 PRIMERS
 Sequencing: TU
 LIBRARY
 Vector : pBAC3.6
 R.Site 1 : EcoRI
 R.Site 2 : EcoRI.
 Location/Qualifiers
 1..1459
 /organism="Pan troglodytes"
 /mol_type="genomic DNA"
 /db_xref="taxon:9598"
 /clone="RP43-008124.TU"
 /sex="male"
 /cell_type="lymphocytes"
 /clone_lib="RP43-008124 Chimpanzee Male BAC Library"
 ORIGIN
 Alignment Scores:
 Pred. No.: 2.36e+04 Length: 1459
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.20% Indels: 0
 DB: 29 Gaps: 0
 US-10-029-345A-109 (1-665) x AG146840 (1-1459)
 QY 486 SerValArgThrSerSerSercJy 493
 |||||
 Db 387 AGCGTCCGACGTCGTGAGTGGG 410
 RESULT 15
 CNSLTT18J 1465 bp mRNA linear HTC 18-JUN-2003
 LOCUS CNSLTT18J/c
 DEFINITION human full-length cDNA clone CS0DK001YK16 of Hela cells of Homo sapiens (human).
 ACCESSION BX248296
 VERSION BX248296.1 GI:28207940
 KEYWORDS HTC.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 1465)
 Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
 Full-length cDNA libraries and normalization
 Unpublished
 Contact: Feng Liang Email: fliang@lifetech.com URL: http://fulllength.invitrogen.com/InvitrogenCorporation 1600 Faraday Avenue
 2 (bases 1 to 1465)
 Genoscope.
 Direct Submission
 Submitted (04-FEB-2003) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail: seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 1st strand cDNA was primed with a NotI-oligo(4T) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by life technologies, a division of invitrogen.
 location/Qualifiers
 1..1465
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0DK001YK16"
 /tissue_type="Hela cells"

CDS

/note="Cot 25-normalized-vector pCMVSPORT_6"
 <1..460
 /note="unnamed protein product"
 /codon_start=2
 /protein_id="CAD62624.1"
 /db_xref="GI:28207941"
 /translation="SRLAGSGFRAPFPFRPPAPKVDAGLEAQRNPRLSVAPLSPV
 PKDYAPORQCPQSNLSLKYMLSGSERVPEPALKSEAVSPILQRIQSRNSIMEA
 AGGFQYHFHNSNDGRFIRTSADLEVISKVSKEVTEKCFIKCTVLF"

ORIGIN

Alignment Scores:
 Pred. No.: 2.37e+04 Length: 1465
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.20% Indels: 0
 DB: 11 Gaps: 0

US-10-029-345A-109 (1-665) x CNSLRT18J (1-1465)

QY 415 AlaserleuHiglyPheSerSer 422
 Db 723 GCCTCTCTACATGCTTTTCACG 700

RESULT 16
 BG121676/c 1502 bp mRNA linear EST 30-JAN-2001
 LOCUS 602352445F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:445067 5',
 DEFINITION mRNA sequence.

ACCESSION BG121676
 VERSION BG121676.1 GI:12615185
 KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1502)
 NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strauberg, Ph.D.
 Email: cgaabs-remail.nih.gov
 Tissue Procurement: ATCC

CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNase Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:

http://image.lnl.gov
 Plate: LLM10236 row: j column: 12
 High quality sequence stop: 65.

Location/Qualifiers
 1..1502

FEATURES

source

/organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:445067"
 /issue_type="adenocarcinoma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /clone_1lb="NIH MGC 90"
 /note="Organ: liver; Vector: pCMV-SPORT6; Site 1: NotI;
 Site 2: SalI; Cloned unidirectionally; oligo-dT primed.
 Average insert size 1.7 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:
 Pred. No.: 2.43e+04 Length: 1502
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0

Query Match: 1.20% Indels: 0
 DB: 12 Gaps: 0

US-10-029-345A-109 (1-665) x BG121676 (1-1502)

QY 383 LeuH16leuSerAlaapArgLeu 390
 Db 1343 CTCGATCTGTCGCGTCATCGTCTT 1320

RESULT 17

BZ578841/c 1603 bp DNA linear GSS 17-DEC-2002
 LOCUS meh2.603.x1 meh Pseudomonas aeruginosa genomic clone meh2_603,
 DEFINITION genomic survey sequence.

ACCESSION BZ578841
 VERSION BZ578841.1 GI:27213902
 KEYWORDS GSS.

SOURCE

ORGANISM Pseudomonas aeruginosa
 Pseudomonas aeruginosa
 Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 Pseudomonadaceae; Pseudomonas.

REFERENCE 1 (bases 1 to 1603)
 Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,
 Burns, J.L., Kaul, R. and Olsen, M.V.
 Whole-Genome-Sequence variation among multiple isolates of
 Pseudomonas aeruginosa library

J. Bacteriol. (2002) In press
 Contact: Chris K. Raymond

COMMENT
 Genome Center
 University of Washington
 Box 352145, Seattle, WA 98105-2145, USA
 Tel: 2062216954
 Fax: 2066857244
 Email: ckraymond@u.washington.edu
 Class: shotgun.

FEATURES

source

Location/Qualifiers
 1..1603
 /organism="Pseudomonas aeruginosa"
 /mol_type="genomic DNA"
 /strain="WSH"
 /db_xref="taxon:287"
 /clone="meh2_603"
 /clone_1lb="msh"
 /note="Environmental isolate. Whole genomic shotgun
 library."

ORIGIN

Alignment Scores:
 Pred. No.: 2.61e+04 Length: 1603
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.20% Indels: 0
 DB: 28 Gaps: 0

US-10-029-345A-109 (1-665) x BZ578841 (1-1603)

QY 7 GlyThrGlnIleValThrGluArg 14
 Db 1387 GGAACCTCAATTTGTACAGGCGA 1364

RESULT 18

AY401953/c 1626 bp DNA linear GSS 12-DEC-2003
 LOCUS Pan troglodytes HCM1062 gene, VIRTUAL TRANSCRIPT, partial sequence,
 DEFINITION genomic survey sequence.

ACCESSION AY401953
 VERSION AY401953.1 GI:39757939
 KEYWORDS GSS.

SOURCE

ORGANISM Pan troglodytes (chimpanzee)
 Pan troglodytes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.

REFERENCE 1 (bases 1 to 1626)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejaritwal,A.,
Todd,M.A., Tanenbaum,D.M., Civeille,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Slnsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Inferring nonneutral evolution from human-chimp-mouse orthologous
gene trios
JOURNAL Science 302 (5652), 1960-1963 (2003)
PUBMED 14671302
REFERENCE 2 (bases 1 to 1626)
AUTHORS Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejaritwal,A.,
Todd,M.A., Tanenbaum,D.M., Civeille,D.R., Lu,F., Murphy,B.,
Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Slnsky,J.J.,
Adams,M.D. and Cargill,M.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
Rockville, MD 20850, USA
COMMENT This sequence was made by sequencing genomic exons and ordering
them based on alignment.
FEATURES
source location/Qualifiers
1..1626
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
<1..>1626
/locus_tag="HCM1062"
ORIGIN
gene
Alignment Scores:
Pred. No.: 2.65e+04 Length: 1626
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 29 Gaps: 0
US-10-029-345A-109 (1-665) x AY401953 (1-1626)
QY 646 LeuGLYLeVAlGYSerGInser 653
Db 326 CTGGGTAAAGTGGGGTCAACATCC 303
RESULT 19 1665 bp mRNA linear EST 21-DEC-2000
BF680537/c 60215527221 NIH_MGC_83 Homo sapiens CDNA clone IMAGE:4296425 5',
LOCUS mRNA sequence.
DEFINITION BF680537.1 GI:11954432
ACCESSION BF680537
VERSION BF680537.1 GI:11954432
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 1665)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-r@mail.nih.gov
Tissue Procurement: CLONETECH Laboratories, Inc.
CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMN at:
http://image.llnl.gov
Plate: LMC1148 row: 0 column: 18
High quality sequence stop: 440.
location/Qualifiers
1..1665
/organism="Homo sapiens"
/mol_type="mRNA"

/db_xref="taxon:9606"
/clone="IMAGE:4296425"
/lab_host="DH10B (T1 phage-resistant)"
/clone_1ib="NIH MGC 83"
/note="Organ: prostate; Vector: pDNR-LIB (Clontech);
Site.1: SfiI (ggcgctcgcc); Site.2: SfiI
(ggcgctcgcc); 5' and 3' adaptors were used in cloning
as follows: 5' adaptor sequence: 5'-ATTCTAGAGCGCGCGCGCATG-dt(30)BN-3' (where B = A,
C, or G and N = A, C, G, or T). Average insert size 1.4
kb (range 0.5-4.0 kb). 14/15 colonies contained inserts
by PCR. This library was enriched for full-length clones
and was constructed by Clontech Laboratories (Palo Alto,
CA)."
ORIGIN
Alignment Scores:
Pred. No.: 2.72e+04 Length: 1665
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 10 Gaps: 0
US-10-029-345A-109 (1-665) x BF680537 (1-1665)
QY 498 SerLeuSerProLeuHisArg 505
Db 1156 TCTGTCTCTCTCCCTTCAACGT 1133
RESULT 20 1684 bp mRNA linear HTC 17-OCT-2002
AY110929 Zea mays CL2242_8 mRNA sequence.
LOCUS AY110929
DEFINITION Zea mays
ACCESSION AY110929
VERSION AY110929.1 GI:21215519
KEYWORDS HTC.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 1684)
AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of
Overgo Probes
JOURNAL Unpublished (2002)
REFERENCE 2 (bases 1 to 1684)
AUTHORS Coe,E.H.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of
Missouri, Columbia, MO 65211, USA
COMMENT If you are interested in getting corresponding physical clones,
these are publicly available from ZMDB and may be found by BLAST
searching at MSU, maizegap.org; ZMDB, www.zmdb.iastate.edu; TIGR,
www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
maize cDNA sequences is either Virginia Walbot, Stanford or Pat
Schubler, Iowa State, then clones may be requested from ZMDB:
www.zmdb.iastate.edu.
location/Qualifiers
1..1684
/organism="Zea mays"
/mol_type="mRNA"
/db_xref="MaizeDB:630884"
/db_xref="taxon:4577"
/clone_1ib="Maize Mapping Project/DuPont Consensus
Library"
/note="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed DuPont contigs; this resource was
assembled by DuPont as part of a collaboration for the

ORIGIN
overgo addressing of BACs in conjunction with the Maize Mapping Project"

Alignment Scores:
Pred. No.: 2.75e+04 Length: 1684
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 11 Gaps: 0

US-10-029-345A-109 (1-665) x AY110929 (1-1684)

QY 549 SerThrpSerLeuThrpSerSer 556
427 TCGACTCCGCTTGACGCTGCC 450

RESULT 21
AK081315 1697 bp mRNA linear HTC 20-SEP-2003
LOCUS AK081315/c
DEFINITION Mus musculus 16 days embryo head cDNA, RIKEN full-length enriched library, clone:CI30003P19 product:unclassified, full insert sequence.

ACCESSION AK081315
VERSION AK081315.1 GI:26099834
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE 1
Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)

JOURNAL MEDLINE
PUBMED 99279253
10349636

REFERENCE

TITLE 2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL MEDLINE
PUBMED 20499374
11042159

REFERENCE

TITLE 3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Kono, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujisake, S., Inoue, K., Togawa, Y., Izawa, M., Omata, E., Watahiki, M.,
Toneida, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL MEDLINE
PUBMED 20530913
11076861

REFERENCE

TITLE 4
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

JOURNAL MEDLINE
PUBMED 11076861

REFERENCE

TITLE 5
The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
6 (bases 1 to 1697)
Adachi, J., Aizawa, K., Akimura, T., Arikawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashide, K., Hayatsu, N., Hiramoto, K., Hirakawa, T., Hirozane, T.,

Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohnato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
Direct Submission

TITLE
JOURNAL
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

COMMENT
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in Riken.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/

FEATURES
source
Location/Qualifiers
1..1697
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM DB:CI30003P19"
/db_xref="MGI:241366"
/db_xref="taxon:10090"
/clone="CI30003P19"
/issue_type="head"
/clone_1b="RIKEN full-length enriched mouse cDNA library"
/dev_stage="16 days embryo"
1..1697
/note="unclassified"

misc_feature

ORIGIN

Alignment Scores:
Pred. No.: 2.77e+04 Length: 1697
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 11 Gaps: 0

US-10-029-345A-109 (1-665) x AK081315 (1-1697)

QY 519 PheGlyLeuSerThrpSerGlnGln 526
1158 TTGGCCTGTCAACGACCAACAA 1135

RESULT 22

LOCUS BE729007 1709 bp mRNA linear EST 15-SEP-2000
DEFINITION 601562269F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:3832024 5',
mRNA sequence.

ACCESSION BE729007
VERSION BE729007.1 GI:10142999

KEYWORDS EST.

SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1709)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-r@mail.nih.gov

Tissue Procurement: ATCC/DCTD/DRP
CDNA Library Prepared by: Ling Hong/Rubin Laboratory
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at: image.linl.gov
plate: L1C4M508 row: 1 column: 17
High quality sequence stop: 30.

FEATURES

source

Location/Qualifiers
1..11709
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3832024"
/cissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_20"
/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCAACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

Alignment Scores:

Pred. No.: 2.79e+04 Length: 1709
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 10 Gaps: 0

US-10-029-345A-109 (1-665) x BE729007 (1-1709)

QY 644 GIJGJLJLJGJLYSValGlySer 651

Db 216 GAGGATTGAGAAAGTGGCTTCC 239

RESULT 23 BZ576364 1730 bp DNA linear GSS 17-DEC-2002
BZ576364LOCUS msh2.4896.v2 msh Pseudomonas aeruginosa genomic clone msh2_4896,
DEFINITION genomic survey sequence.ACCESSION BZ576364
VERSION BZ576364.1 GI:27211425

KEYWORDS

GSS.

SOURCE

Pseudomonas aeruginosa

ORGANISM

Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

REFERENCE

1 (bases 1 to 1730)

AUTHORS

Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,

TITLE

Whole-Genome-Sequence variation among multiple isolates of

JOURNAL

Pseudomonas aeruginosa library

COMMENT

U. Bacteriol. (2002) In press

CONTACT

Contact: Chris K. Raymond

UNIVERSITY

Genome Center

ADDRESS

University of Washington

CITY

Box 352145, Seattle, WA 98105-2145, USA

TEL

Tel: 2062216954

FAX

Fax: 2066857244

EMAIL

Email: craymond@u.washington.edu

CLAS

Class: shotgun.

FEATURES

Location/Qualifiers

source

1..1730

organism

/organism="Pseudomonas aeruginosa"

mol_type

/mol_type="genomic DNA"

strain

/strain="msh"

db_xref

/db_xref="taxon:287"

/clone="msh2_4896"
/clone_lib="msh"
/note="Environmental isolate. Whole genomic shotgun library."

ORIGIN

Alignment Scores:

Pred. No.: 2.83e+04 Length: 1730
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 28 Gaps: 0

US-10-029-345A-109 (1-665) x BZ576364 (1-1730)

QY 580 SerAlaTYSerCYSerGlnLeu 587

Db 372 TCCGGTACTCATGCTCACGCTG 395

RESULT 24

BG859060

LOCUS

DEFINITION

1024061B08.x1 C. reinhardtii CC-1690, normalized, Lambda Zap II

ACCESSION

BG859060.1 GI:14240244

KEYWORDS

EST.

SOURCE

Chlamydomonas reinhardtii

ORGANISM

Chlamydomonas reinhardtii
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

REFERENCE

1 (bases 1 to 1742)

AUTHORS

Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P.,

TITLE

Mcdermott,J.P., Silflow,C., Stern,D. and Surzycki,R.
Analyses of the Chlamydomonas reinhardtii Genome: A Model,
Unicellular System for Analyzing Gene Function and Regulation in

JOURNAL

Vascular Plants: project phase 2

COMMENT

Unpublished (2000)

CONTACT

Contact: Charles Hauser

UNIVERSITY

Duke University

CITY

Durham, NC 27708-1000

TEL

Tel: 919 613 8159

FAX

Fax: 919 613 8177

EMAIL

Email: chauser@duke.edu.

FEATURES

Location/Qualifiers

source

1..1742

organism

/organism="Chlamydomonas reinhardtii"

mol_type

/mol_type="mRNA"

strain

/strain="CC-1690 wild type mt+ 21gr"

db_xref

/db_xref="taxon:3055"

clone_lib

/clone_lib="C. reinhardtii CC-1690, normalized, Lambda Zap II"

note

/note="Vector: pBluescript II SK-; Site 1: EcoRI; Site 2: XhoI; This library, constructed by John Davies and Jeffrey

Mcdermott

Mcdermott, combines cDNAs from CC-1690 cells grown to

mid-log phase

in TAP (acetate-containing) medium in the

light

light, TAP medium in the dark, HS (minimal) medium in

ambient levels

of CO2 and HS medium bubbled with 5% CO2.

POLYA

mRNA was purified from each sample, pooled and cDNA

synthesized

The cDNA was directionally cloned into lambda

ZAP II

(Stratagene) in the EcoRI (5') and XhoI (3') sites.

pBluescript

II SK- plasmids were excised from the lambda

phage

ZAP clones by superinfection with ExSist (Stratagene)

phage

The library was normalized using method 4 described

in

Bonaldo et al (1996) Genome Research 6: 791-806."

ORIGIN

Alignment Scores:

Pred. No.: 2.85e+04 Length: 1742
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.20% Indels: 0
 DB: 12 Gaps: 0

US-10-029-345A-109 (1-665) x BG859060 (1-1742)

Qy 473 ThralaarpSeraeSergin 480
 |||||
 Db 106 ACCGACGCCCGACGACCTCTCAG 129

RESULT 25
 AY400863 1746 bp DNA linear GSS 12-DEC-2003
 LOCUS Homo sapiens HCM0699 gene, VIRTUAL TRANSCRIPT, partial sequence,
 DEFINITION genomic survey sequence.
 ACCESSION AY400863.1 GI:39756852
 VERSION AY400863.1
 KEYWORDS GSS.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1746)
 Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
 Todd,M.A., Tanenbaum,D.M., Civeello,D.R., Lu,F., Murphy,B.,
 Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Snihsy,J.J.,
 Adams,M.D. and Cargill,M.
 Direct Submission
 Interferring nonneutral evolution from human-chimp-mouse orthologous
 gene clis08
 JOURNAL Science 302 (5652), 1960-1963 (2003)
 PUBMED 14671302
 REFERENCE 2 (bases 1 to 1746)
 Clark,A.G., Glanowski,S., Nielson,R., Thomas,P., Kejarawal,A.,
 Todd,M.A., Tanenbaum,D.M., Civeello,D.R., Lu,F., Murphy,B.,
 Ferreira,S., Wang,G., Zheng,X.H., White,T.J., Snihsy,J.J.,
 Adams,M.D. and Cargill,M.
 Direct Submission
 Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive,
 Rockville, MD 20850, USA
 COMMENT This sequence was made by sequencing genomic exons and ordering
 them based on alignment.
 FEATURES
 source
 1. 1746
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 <1. 21746
 /locus_tag="HCM0699"

ORIGIN
 Alignment Scores:
 Pred. No.: 2.86e+04 Length: 1746
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.20% Indels: 0
 DB: 29 Gaps: 0

US-10-029-345A-109 (1-665) x AY400863 (1-1746)

Qy 498 SerleulenseProleuHsArcy 505
 |||||
 Db 687 AGCCTCTTTCACCACTTCATCGC 710

RESULT 26
 BM320908 1770 bp mRNA linear EST 03-JAN-2002
 LOCUS rockefeller.0.435 Mastigamoeba balamutchi lambda ZAP II library
 DEFINITION Mastigamoeba balamutchi cDNA similar to actin-binding protein
 (coronin), mRNA sequence.
 ACCESSION BM320908
 VERSION BM320908.1 GI:18055314
 KEYWORDS EST.

SOURCE Mastigamoeba balamutchi
 ORGANISM Mastigamoeba balamutchi
 Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
 REFERENCE 1 (bases 1 to 1770)
 Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
 Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
 Philippe,H.
 TITLE The analysis of 100 genes supports the grouping of three highly
 divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
 MEDLINE 21819461
 PUBMED 11830664
 COMMENT Contact: Muller Miklos
 Laboratory of Biochemical Parasitology
 The Rockefeller University
 1230 York Avenue, New York, NY 10021, USA
 Email: mmuller@rockvax.rockefeller.edu
 Insert Length: 1770 Std Error: 0.00
 POLYA=Yes.

FEATURES
 source
 1. 1770
 /organism="Mastigamoeba balamutchi"
 /mol_type="mRNA"
 /strain="ATCC 30984"
 /db_xref="taxon:108607"
 /clone_1ib="Mastigamoeba balamutchi lambda ZAP II library"
 /note="syn: Phreatamoeba balamutchi"

ORIGIN
 Alignment Scores:
 Pred. No.: 2.9e+04 Length: 1770
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.20% Indels: 0
 DB: 12 Gaps: 0

US-10-029-345A-109 (1-665) x BM320908 (1-1770)

Qy 488 ArgThSerserSerg1YThAla 495
 |||||
 Db 1324 AGAACGACGACCTCAGAACCCGC 1347

RESULT 27
 BM321015 1792 bp mRNA linear EST 03-JAN-2002
 LOCUS rockefeller.0.586 Mastigamoeba balamutchi lambda ZAP II library
 DEFINITION Mastigamoeba balamutchi cDNA similar to actin-binding protein
 (coronin), mRNA sequence.
 ACCESSION BM321015
 VERSION BM321015.1 GI:18055421
 KEYWORDS EST.
 SOURCE Mastigamoeba balamutchi
 ORGANISM Mastigamoeba balamutchi
 Eukaryota; Pelobiontida; Mastigamoebidae; Mastigamoeba.
 REFERENCE 1 (bases 1 to 1792)
 Bapteste,E., Brinkmann,H., Lee,J.A., Moore,D.V., Sensen,C.W.,
 Gordon,P., Durufle,L., Gaasterland,T., Lopez,P., Muller,M. and
 Philippe,H.
 TITLE The analysis of 100 genes supports the grouping of three highly
 divergent amoebae: Dictyostelium, Entamoeba, and Mastigamoeba
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (3), 1414-1419 (2002)
 MEDLINE 21819461
 PUBMED 11830664
 COMMENT Contact: Muller Miklos
 Laboratory of Biochemical Parasitology
 The Rockefeller University
 1230 York Avenue, New York, NY 10021, USA
 Email: mmuller@rockvax.rockefeller.edu
 Insert Length: 1792 Std Error: 0.00
 POLYA=Yes.

FEATURES
 source
 1. 1792
 /organism="Mastigamoeba balamutchi"

```

/organism="Mus musculus"
/mol_type="cDNA"
/strain="ATCC 30984"
/db_xref="taxon:10607"
/clone_id="Mus musculus balaamuthi lambda ZAP II library"
/notes="syn: Phreataeoba balaamuthi"

ORIGIN

Alignment Scores:
Pred. No.: 2.94e+04 Length: 1792
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: Gaps: 0

US-10-029-345A-109 (1-665) x BM321015 (1-1792)

QY 488 ArgTherSergGlyThrAla 495
Db 1346 AGAAGCAGAGCTCAGACCGG 1369

RESULT 28
AC045836 1839 bp mRNA linear HTC 20-SEP-2003
LOCUS Mus musculus adult male corpora quadrigemina cDNA, RIKEN
DEFINITION full-length enriched library, clone: B230313L14 product: unknown EST,
full insert sequence.
ACCESSION AK045836 GI:26091148
VERSION AK045836.1
KEYWORDS HTC, CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636

AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159

AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujitake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861

AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the
PANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
AUTHORS 5
TITLE The PANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
JOURNAL Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

```

```

REFERENCE
AUTHORS 6 (bases 1 to 1839)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T.,
Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
Kato, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shitaku, T.,
Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
Muramatsu, M. and Hayashizaki, Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Shiohira-cho, Tsunumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.
FEATURES
source location/Qualifiers
1..1839
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:B230313L14"
/db_xref="MG1:2411215"
/db_xref="taxon:10090"
/clone="B230313L14"
/sex="male"
/tissue_type="corpora quadrigemina"
/clone_id="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
1..1839
/notes="unknown EST (GB|AA689954, evidence: BLASTN, 99%,
match=358)"

misc_feature
misc_feature

ORIGIN

Alignment Scores:
Pred. No.: 3.02e+04 Length: 1839
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: Gaps: 0

US-10-029-345A-109 (1-665) x AK045836 (1-1839)

QY 171 GlnArgAspValLeuGlnGlu 178
Db 275 CAGCGAGATGTCTGAAAGAA 298

RESULT 29
AC017699 1867 bp mRNA linear HTC 20-SEP-2003
LOCUS Mus musculus 8 days embryo whole body cDNA, RIKEN full-length
DEFINITION enriched library, clone: 573047606 product: D11069P2.2 (OUTER
MITOCHONDRIAL MEMBRANE TRANSLOCASE HTOM34P) (SIMILAR TO TRANSLOCASE
OF OUTER MITOCHONDRIAL MEMBRANE 34) (HYPOTHETICAL 34.6 KDa PROTEIN)
(UNKNOWN) (PROTEIN FOR MGC:3515) homolog (Homo sapiens), full
insert sequence.
ACCESSION AK017699
VERSION AK017699.1 GI:12857072
KEYWORDS HTC, CAP trapper.

```


SOURCE ORGANSIM	Mus musculus (house mouse)
REFERENCE AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
JOURNAL MEDLINE PUBMED	Carninci, P. and Hayashizaki, Y. High-efficiency long-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999) 99279253 10349656
REFERENCE AUTHORS	2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komo, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000) 20499374 11042159
JOURNAL MEDLINE PUBMED	3 Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Saeki, N., Carninci, P., Komo, H., Akiyama, Y., Nishi, K., Kitsumata, T., Tashiro, H., Itoh, M., Suni, N., Ishii, Y., Nakamura, S., Hazama, N., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasaiwagi, K., Fujiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsubiki, M., Yoneda, Y., Ichikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kiria, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000) 20530913 11076661
JOURNAL MEDLINE PUBMED	4 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium. Functional annotation of a full-length mouse cDNA collection Nature 409, 685-690 (2001) 5 The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002) 6 (bases 1 to 1867) Adachi, J., Aizawa, K., Akabira, S., Aktamura, T., Arai, A., Aono, H., Furukawa, T., Bono, H., Carninci, P., Fukuda, S., Fukushima, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hirotsuka, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kaushika, T., Kato, H., Kawai, J., Kojima, Y., Kono, H., Kouda, M., Koys, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Salto, H., Salto, R., Sakai, C., Sakai, K., Sano, H., Saeki, D., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Suuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toyota, T., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M. and Hayashizaki, Y. Direct Submission Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Physical and Chemical Research (RIKEN), Laboratory for Genome Physical and Chemical Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan [E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216] Please visit our web site (http://genome.gsc.riken.go.jp/) for further details.
TITLE JOURNAL	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5' GAGACGACAAAGATCCAGAGCTTTTTTTTTTTTTTN 3'], cDNA was prepared by using triethanol thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went
COMMENT	

through one round of subreaction to Rot = 100.0. Second strand cDNA was prepared with the primer adaptor of sequence [5' GAGAGACATTCGACGTTATTAATTAATCCCCCCCCC 3']. cDNA was cleaved with BamI and XhoI. Vector: a modified pluscript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI, 3' end: BamI. Host: DH10B.

Location/Qualifiers

1..1867

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="FANTOM DB:5730476E06"

/db_xref="MGJ:1838741"

/db_xref="taxon:10090"

/clone="5730476E06"

/tissue_type="whole body"

/clone_lib="RIKEN full-length enriched mouse cDNA library"

/dev_stage="8 days embryo"

76..1005

/note="unnamed protein product; DJ1069P2.2 (OUTER MITOCHONDRIAL MEMBRANE TRANSLOCASE HTOM34P) (SIMILAR TO TRANSLOCASE OF OUTER MITOCHONDRIAL MEMBRANE 34) (HYPOTHETICAL 34.6 KDA PROTEIN) (UNKNOWN) (PROTEIN FOR MGC:3515) homolog (Homo sapiens) [SPR109NT23, evidence: PASTY, 87.1%ID, 100%length, match=927]

putative"

/codon_start=1

/protein_id="BAB30882.1"

/db_xref="GI:12657073"

/translation="MAPKVSDFVEQLRAGNQNPRNGVGEASALYERALLQAGSSADDEESVLYSNRACTLDGNGCTDCKTSLALVPYSIKRLPRASAYELKETA LAVDYKTVLQIDNSVASALEGINRITRALMDSLGPEWRKLPPIVPVPSAQKRNLSLPDNHETAKTKTSKCATLTKSRVPSGADVERAKLKEBENDLVKGNHKA1EKSE SLICSSLESAATYENRALCHLVLRQYEAVDCTEALKLDSKNVKAFFRRAQAYKALDYKSSLDISLSLQIEPRNGAQKLRQEVNQNNA"

1848..1853

/note="putative"

1867

/note="putative"

polya_signal

polya_site

ORIGIN

Alignment Scores:

Pred. NO.:

Score: 3.07e+04 Length: 1867

8.00 Matches: 8

100.00% Conservatvie: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 1.20% Indels: 0

DB: 11 Gaps: 0

US-10-029-345A-109 (1-665) x AK017699 (1-1867)

Qy 384 HisLenSerAlaApArgLenGu 391

Db 1421 CACCTGAGTGTGACAGGCTTAGAA 1444

RESULT 30

AK075873

LOCUS

DEFINITION Mus musculus adult male tongue cDNA, Riken full-length enriched library, clone:2310093L17 product:DJ1069P2.2 (OUTER MITOCHONDRIAL MEMBRANE TRANSLOCASE HTOM34P) (SIMILAR TO TRANSLOCASE OF OUTER MITOCHONDRIAL MEMBRANE 34) (HYPOTHETICAL 34.6 KDA PROTEIN) (UNKNOWN) (PROTEIN FOR MGC:3515) homolog (Homo sapiens), full insert sequence.

AK075873

ACCESSION AK075873

VERSION AK075873.1

GI:26344741

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 Carninci, P. and Hayashizaki, Y.

AUTHORS

TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Katsunuma, T., Tashiro, H., Itoh, M., Sumi, N., Iehi, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the PANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
MEDLINE 11076861
PUBMED 11076861
REFERENCE 5
AUTHORS The PANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
MEDLINE 11076861
PUBMED 11076861
REFERENCE 6
AUTHORS Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Aono, H., Arai, A., Arahawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y., Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K., Hirao, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kono, H., Kozawa, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Okazaki, Y., Okido, T., Owa, C., Saito, H., Saito, R., Sakai, C., Sakai, K., Sano, H., Sasaki, D., Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagami, M., Tagawa, A., Takahashi, F., Tanaka, T., Tejima, Y., Toya, T., Yamamura, T., Yamana, I., Yasunishi, A., Yoshida, K., Yoshino, M., Muramatsu, M., and Hayashizaki, Y.
TITLE Direct Submission
JOURNAL Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 220-0045, Japan (E-mail: genome-res@gs.c.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
COMMENT CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/
Location/Qualifiers
1. 1918
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CS7BL/6J"
/db_xref="PANTOM DB:2310009L17"
/db_xref="MGI:1895159"
/db_xref="taxon:10090"

CDS
/clone="2310009L17"
/sex="male"
/tissue_type="longue"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
128-1057
/note="unnamed protein product; D11069P2.2 (OUTER MITOCHONDRIAL MEMBRANE TRANSLOCASE HTOM34P) (SIMILAR TO TRANSLOCASE OF OUTER MITOCHONDRIAL MEMBRANE 34) (HYPOTHETICAL 34.6 KDA PROTEIN) (UNKNOWN) (PROTEIN FOR MGC:3515) homolog [Homo sapiens] (SPTR|Q9NTZ3, evidence: PASTY, 87.1%ID, 100%length, match=927)
putative"
/codon_start=1
/protein_id="BAC36020.1"
/db_xref="GI:26344742"
/translation="MAPVSPSVQLRAAGNFRNGOYGEASALYERALTILQARSLADPEESLYENRACVLDKNCCTDKDTSALATVPSTKPLLRASAYEAEKXALAYDYKVLQIDNSVASALEGIRITRALMDSIGPEWRKLPPIVPSAQRKMSLPDSNKEATKSKREKATKSRVPSGADVERALKEBNDLVKKNHKKALKESYSLCSLSLEATYSNRALCHLVKQKAVDCTALKLDGNVAFYRRAQYALMDYKSSLSIDISLQIEPRNGPQKLRQVNMNM"
1901. 1906
/note="putative"
1918
/note="putative"
polya_signal
polya_site
ORIGIN
Alignment Scores:
Pred. No.: 3,16e+04 Length: 1918
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatch: 0
Query Match: 1.20% Indels: 0
DB: 11 Gaps: 0
US-10-029-345A-109 (1-665) x AK075873 (1-1918)
QY 384 HistoneSer1aaPpAglenglu 391
Db 1474 CACCTGAGTGTGACAGCTAGAA 1497
RESULT 31
CG755756/c 1919 bp DNA linear GSS 24-OCT-2003
LOCUS CG755756
DEFINITION CG755756, genomic survey sequence.
ACCESSION CG755756
VERSION CG755756.1 GI:37982573
KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida; Nododiplogasteridae; Pristionchus.
1 (bases 1 to 1919)
Srinivasan, J., Srin, W., Jesse, T., Wiggers-Perebolte, L., Jansen, K., Buntjer, J., van der Meulen, M., and Sommer, R. U.
An integrated physical and genetic map of the nematode Pristionchus pacificus
Mol. Genet. Genomics 269 (5), 715-722 (2003)
22835951
12884007
Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de
Class: BAC ends.
Location/Qualifiers
1. 1919
/organism="Pristionchus pacificus"

```

/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Epa EcoRI BAC Library"
/note="The library was generated by a partial digest of
the genomic DNA with EcoRI and cloning into the BAC
vector."

```

ORIGIN

Alignment Scores:

Pred. No.:	3.16e+04	Length:	1919
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.20%	Indels:	0
DB:	29	Gaps:	0

US-10-029-345A-109 (1-665) x CG755756 (1-1919)

QY 548 ThrSerThrProSerLeuThrSer 555

Db 29 ACCAGCAGCACCTCCCTGACATCA 6

RESULT 32

LOCUS AK047699/c

AK047699 1965 bp mRNA linear HTC 20-SEP-2003
 Mus musculus adult male corpus striatum cDNA, RIKEN full-length
 enriched library, clone: C030013H07 product: unclassifiable, full
 insert sequence.

ACCESSION

AK047699

VERSION

AK047699.1 GI:26092361

KEYWORDS

HTC, CAP trapper.

SOURCE

Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

AUTHORS

1 Carninci, P. and Hayashizaki, Y.

TITLE

High-efficiency full-length cDNA cloning

JOURNAL

Meth. Enzymol. 303, 19-44 (1999)

MEDLINE

PUBMED

AUTHORS

REFERENCE

PUBMED

AUTHORS

TITLE

1 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL

Genome Res. 10 (11), 1757-1771 (2000)

MEDLINE

PUBMED

AUTHORS

TITLE

4 The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

JOURNAL

MEDLINE

PUBMED

AUTHORS

TITLE

5 The FANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs

JOURNAL Nature 420, 563-573 (2002)

REFERENCE

AUTHORS

6 (bases 1 to 1965)
 Adachi, J., Aizawa, K., Akimura, T., Arikawa, T., Bono, H., Carninci, P.,
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hattizume, M.,
 Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoaka, T., Hirozane, T.,
 Hori, F., Ito, K., Itoh, M., Itoh, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
 Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
 Koyama, S., Kurahara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishii, K., Nomura, K., Nunazaki, R., Ohno, M., Ohsato, N.,
 Okazaki, Y., Saito, R., Saitoh, K., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
 Muramatsu, M. and Hayashizaki, Y.

TITLE

JOURNAL

SUBMITTED

DIRECT

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

SUBMITTED

REFERENCE 1 (bases 1 to 2008)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
JOURNAL Pseudomonas aeruginosa library
COMMENT J. Bacteriol. (2002) in press
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
Location/Qualifiers
1..2008
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pac2-164.4936"
/clone_1lb="pac2-164"
/note="clinical isolate 2-164 whole genomic shotgun
library."

Alignment Scores:
Pred. No.: 3.32e+04 Length: 2008
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 28 Gaps: 0

US-10-029-345A-109 (1-665) x BZ564814 (1-2008)

QY 357 ProAlaServAlProServAlPro 364
1372 CCGGCTCTGTCTCCCTCGGCGCN 1349

RESULT 34
BF572336/c 2174 bp mRNA linear EST 12-DEC-2000
LOCUS 60207782FI NIH_MGC_62 Homo sapiens CDNA IMAGE:4252004 5',
DEFINITION mRNA sequence.
ACCESSION BF572336 2174 bp mRNA linear EST 12-DEC-2000
VERSION BF572336 GI:11646036
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 2174)
REFERENCE NIH-MGC http://mgi.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCMD/DTF
CDNA Library Preparation: CLONTECH Laboratories, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LNLN)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov
Plate: LNCM1076 row: d column: 21
High quality sequence start: 4
High quality sequence stop: 26.
Location/Qualifiers
1..2174
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"

FEATURES
source

/clone="IMAGE:4252004"
/tissue_type="melanotic melanoma, high MDR"
/lab_host="DH10B (T1 phage-resistant)"
/clone_1lb="NIH MGC_62"
/note="Organ: skin; Vector: pDNR-LIB (Clontech); Site_1:
SfiI (ggccgctggcc); Site_2: SfiI (ggccatggcc);
Double-stranded cDNA was prepared from cell line RNA. 5'
and 3' adaptors were used in cloning as follows: 5'
adaptor sequence: 5'-CACGGCATTATGGCC-3' and 3' adaptor
sequence: 5'-ATTCTAGAGCGCGCGGCACATG-dt(30)BN-3'
(where B = A, C, or G and N = A, C, G, or T). Average
insert size 1.75 kb (range 0.9-4.0 kb). 15/15 colonies
contained inserts by PCR. This library was enriched for
full-length clones and was constructed by Clontech
Laboratories (Palo Alto, CA)."

ORIGIN

Alignment Scores:

Pred. No.: 3.61e+04 Length: 2174
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 10 Gaps: 0

US-10-029-345A-109 (1-665) x BF572336 (1-2174)

QY 323 ValProAlaValSerGluGly 330
DB 681 GTGCTCTGTGTGCGAGGCGCGG 658

RESULT 35
AW730576/c

LOCUS AW730576 2208 bp mRNA linear EST 08-MAR-2001
DEFINITION GA_Ba0026P07 Gossypium arboreum 7-10 dpa fiber library Gossypium
ARBOREUM CDNA clone GA_Ba0026P07, mRNA sequence.
ACCESSION AW730576
VERSION AW730576 GI:7628222
KEYWORDS EST.
SOURCE Gossypium arboreum
ORGANISM Gossypium arboreum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Malvales; Malvaceae; Malvaceae; Gossypium.
1 (bases 1 to 2208)
REFERENCE Wang,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J.,
Henry,D., Wood,T.C., Leslie,A. and Wilkins,T.A.
TITLE An integrated analysis of the genetics, development, and evolution
JOURNAL of the cotton fiber
COMMENT Unpublished (2000)
Contact: Wang RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rtwing@clemson.edu
High quality sequence stop: 2208.
Location/Qualifiers
1..2208
/organism="Gossypium arboreum"
/mol_type="mRNA"
/strain="AKA"
/cultivar="8400"
/db_xref="taxon:29729"
/clone="GA_Ba0026P07"
/tissue_type="Fibers isolated from bolls harvested 7-10
dpa"
/lab_host="R. coli"
/clone_1lb="Gossypium arboreum 7-10 dpa fiber library"
/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

ORIGIN

US-10-029-345A-109 (1-665) x AK011786 (1-2302)

QY 384 HisLeuSer1AAspArgLeuGlu 391
 |||||
 Db 1855 CACCTGAGTGTGACAGCTCAGAA 1878

RESULT 37
 AY109354/c
 LOCUS AY109354 2472 bp mRNA linear HTC 17-OCT-2002
 DEFINITION Zea mays C1199_1 mRNA sequence.
 ACCESSION AY109354
 VERSION AY109354.1 GI:21213009
 KEYWORDS HTC.
 SOURCE Zea mays
 ORGANISM Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Zea.

REFERENCE
 1 (bases 1 to 2472)
 Hailey,C.F., Dolan,M., Mao,G.H., Vogel,J.M., Whitsitt,M.S.,
 Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
 TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of
 Overgo Probes
 JOURNAL Unpublished (2002)
 REFERENCE 2 (bases 1 to 2472)
 TITLE Direct Submission
 AUTHORS Submitted (25-APR-2002) Maize Mapping Project, University of
 Missouri, Columbia, MO 65211, USA
 COMMENT If you are interested in getting corresponding physical clones,
 these are publicly available from zmdb and may be found by BLAST
 searching at MSL, maizemap.org; ZMDB, www.zmdb.iastate.edu; TIGR,
 www.tigr.org; or NCBI, www.ncbi.nlm.nih.gov. When the source of the
 maize cDNA sequences is either Virginia Walbot, Stanford or Pat
 Schnable, Iowa State, then clones may be requested from ZMDB:
 www.zmdb.iastate.edu.

FEATURES
 source Location/Qualifiers
 1..2472
 /organism="Zea mays"
 /mol_type="mRNA"
 /db_xref="taxon:4577"
 /clone_lib="Maize Mapping Project/DuPont Consensus
 library"
 /note="this sequence is part of a project of EST
 assemblies resulting from the application of public
 contigs to seed DuPont contigs; this resource was
 assembled by DuPont as part of a collaboration for the
 overgo addressing of BACs in conjunction with the Maize
 Mapping Project"

ORIGIN
 Alignment Scores:
 Pred. No.: 4.14e+04 Length: 2472
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.20% Indels: 0
 DB: 11 Gaps: 0

US-10-029-345A-109 (1-665) x AY109354 (1-2472)

QY 486 SerValArgThrSerSerGly 493
 |||||
 Db 2090 TCTGTAAAGACTTCATCTCTGGG 2067

RESULT 38
 AK084572/c
 LOCUS AK084572 2635 bp mRNA linear HTC 20-SEP-2003
 DEFINITION Mus musculus 13 days embryo heart cDNA, RIKEN full-length enriched
 library, clone: D3J0016F04 Product: unknown EST, full insert
 sequence.
 ACCESSION AK084572

VERSION AK084572.1 GI:26102078
 KEYWORDS HTC; CAP trapper
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
 1
 Carninci,P. and Hayashizaki,Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636

AUTHORS
 2
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
 Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499374
 PUBMED 11042159

AUTHORS
 3
 Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
 Kono,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M.,
 Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A.,
 Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
 Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watabiki,M.,
 Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J.,
 Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
 TITLE RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multichipillary sequencer
 JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
 MEDLINE 20530913
 PUBMED 11076861

AUTHORS
 4
 The RIKEN Genome Exploration Research Group Phase II Team and the
 PANTOM Consortium.
 TITLE Functional annotation of a full-length mouse cDNA collection
 JOURNAL Nature 409, 685-690 (2001)
 REFERENCE 5
 The PANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 TITLE Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 JOURNAL Nature 420, 563-573 (2002)
 REFERENCE 6 (bases 1 to 2635)
 Nature 420, 563-573 (2002)

AUTHORS
 Adachi,J., Aizawa,K., Akimura,T., Arikawa,T., Bono,H., Carninci,P.,
 Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
 Hayashida,K., Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T.,
 Hori,F., Imocani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
 Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Kono,H., Kouda,M.,
 Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
 Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ono,M., Ohsato,N.,
 Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
 Sano,H., Sasaki,D., Shibata,K., Shingawa,A., Shikata,T.,
 Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
 Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
 Muramatsu,M. and Hayashizaki,Y.
 TITLE Direct Submission
 JOURNAL Submitted (16-APR-2002) Yoshinide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan [E-mail:genome-res@gsc.riken.go.jp,
 URL:ftp://genome.gsc.riken.go.jp, Tel:81-45-503-9222,
 Fax:81-45-503-9216]
 COMMENT cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in Riken
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Please visit our web site for further details.
 URL:ftp://genome.gsc.riken.go.jp/
 URL:http://fantom.gsc.riken.go.jp/

```

FEATURES
source
    Location/Qualifiers
    1. 2635
       /organism="Mus musculus"
       /mol_type="mRNA"
       /strain="C57BL/6J"
       /db_xref="PANTOM.DB:D330016F04"
       /db_xref="MGI:2431227"
       /db_xref="taxon:10090"
       /clone="D330016F04"
       /issue_type="heart"
       /clone_lib="RIKEN full-length enriched mouse cDNA library"
       /dev_stage="13 days embryo"
       1. 2635
       /note="unknown EST (GB|BB012554, evidence: BLASTN, 99%,
       match=495)"

ORIGIN
Alignment Scores:
Pred. No.:      4.43e+04      Length:      2635
Score:          8.00         Matches:      8
Percent Similarity: 100.00%   Conservative: 0
Beet Local Similarity: 100.00% Mismatches: 0
Query Match:     1.20%       Indels:      0
DB:              11         Gaps:        0

US-10-029-345A-109 (1-665) x AK084572 (1-2635)

QY      526 GlnHleuthrlySeraIagly 533
Db      204 CAACATCTGACCAAGTCAGCTGCG 181

RESULT 39
AK031373
LOCUS   AK031373      2739 bp      mRNA      linear      HTC 18-SEP-2003
DEFINITION Mus musculus 13 days embryo male testis cDNA, RIKEN full-length
            enriched library, clone:6030414M07 product:zinc finger protein 91,
            full insert sequence.
ACCESSION AK031373      GI:26082294
VERSION   AK031373.1
KEYWORDS  HTC; CAP trapper.
SOURCE    Mus musculus (house mouse)
ORGANISM  Mus musculus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS   Carninci,P. and Hayashizaki,Y.
TITLE     High-efficiency full-length cDNA cloning
JOURNAL   Meth. Enzymol. 303, 19-44 (1999)
MEDLINE   99279253
PUBMED    10349636

REFERENCE
AUTHORS   Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
          Itoh,M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE     Normalization and subtraction of cap-trapper-selected cDNAs to
          prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL   Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE   20499374
PUBMED    11042159

REFERENCE
AUTHORS   Shibata,K., Itoh,M., Aizawa,K., Nagaoaka,S., Sasaki,N., Carninci,P.,
          Kono,H., Akiyama,T., Nishi,K., Kiteunai,T., Tashiro,H., Itoh,M.,
          Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A.,
          Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
          Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Wataniki,M.,
          Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsumura,S., Kawai,D.,
          Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE     RIKEN integrated sequence analysis (RISA) system--384-format
          sequencing pipeline with 384 multicapillary sequencer
JOURNAL   Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE   11076861
PUBMED    11076861

AUTHORS   The RIKEN Genome Exploration Research Group Phase II Team and the

```

```

TITLE     FANTOM Consortium.
JOURNAL   Functional annotation of a full-length mouse cDNA collection
REFERENCE Nature 409, 685-690 (2001)
AUTHORS   5

TITLE     The PANTOM Consortium and the RIKEN Genome Exploration Research
          Group Phase I & II Team.
JOURNAL   Analysis of the mouse transcriptome based on functional annotation
REFERENCE of 60,770 full-length cDNAs
AUTHORS   Nature 420, 563-573 (2002)
          6 (bases 1 to 2739)
          Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
          Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoaka,T., Hirozane,T.,
          Kato,H., Kawai,Y., Kojima,Y., Kondo,S., Kono,H., Kouda,M.,
          Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
          Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohata,N.,
          Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
          Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
          Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
          Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
          Muramatsu,M. and Hayashizaki,Y.
          Direct Submision
          Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
          Physical and Chemical Research (RIKEN), Laboratory for Genome
          Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
          Kanagawa 230-0045, Japan (E-mail:genome-res@gsr.riken.go.jp,
          URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
          Fax:81-45-503-9216)

COMMENT   cDNA library was prepared and sequenced in Mouse Genome
          Encyclopedia Project of Genome Exploration Research Group in Riken
          Genomic Sciences Center and Genome Science Laboratory in RIKEN.
          Division of Experimental Animal Research in Riken contributed to
          prepare mouse tissues.
          Please visit our web site for further details.
          URL:http://genome.gsc.riken.go.jp/
          URL:http://fantom.gsc.riken.go.jp/.
          Location/Qualifiers
          1. 2739
             /organism="Mus musculus"
             /mol_type="mRNA"
             /strain="C57BL/6J"
             /db_xref="PANTOM.DB:6030414M07"
             /db_xref="MGI:2394592"
             /db_xref="taxon:10090"
             /clone="6030414M07"
             /sex="male"
             /issue_type="testis"
             /clone_lib="RIKEN full-length enriched mouse cDNA library"
             /dev_stage="13 days embryo"
             1. 2739
             /note="zinc finger protein 91 (locuslink|109910,
             GB|U05342, evidence: BLASTN, 99%, match=838) "

```

```

FEATURES
source
    Location/Qualifiers
    1. 2739
       /organism="Mus musculus"
       /mol_type="mRNA"
       /strain="C57BL/6J"
       /db_xref="PANTOM.DB:6030414M07"
       /db_xref="MGI:2394592"
       /db_xref="taxon:10090"
       /clone="6030414M07"
       /sex="male"
       /issue_type="testis"
       /clone_lib="RIKEN full-length enriched mouse cDNA library"
       /dev_stage="13 days embryo"
       1. 2739
       /note="zinc finger protein 91 (locuslink|109910,
       GB|U05342, evidence: BLASTN, 99%, match=838) "

ORIGIN
Alignment Scores:
Pred. No.:      4.62e+04      Length:      2739
Score:          8.00         Matches:      8
Percent Similarity: 100.00%   Conservative: 0
Beet Local Similarity: 100.00% Mismatches: 0
Query Match:     1.20%       Indels:      0
DB:              11         Gaps:        0

US-10-029-345A-109 (1-665) x AK031373 (1-2739)

QY      414 AaaAaSerleuHieglYpbeser 421
Db      1801 GCAGCATCCCTCCATGCTTCTCC 1824

RESULT 40
AK078652/c
LOCUS   AK078652      3012 bp      mRNA      linear      HTC 18-SEP-2003

```


DEFINITION	Mus musculus adult male adrenal gland cDNA, RIKEN full-length gene 3, full insert sequence.
ACCESSION	AK078652
VERSION	AK078652.1 GI:26098017
KEYWORDS	HTC; CAP trapper.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	Carninci, P. and Hayashizaki, Y.
TITLE	1 High-efficiency full-length cDNA cloning
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)
MEDLINE	99279253
PUBMED	10349636
REFERENCE	2
AUTHORS	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE	3 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE	20493374
PUBMED	11042159
REFERENCE	3
AUTHORS	Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P., Komno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Suna, I., Ishii, Y., Nakamura, S., Hazama, N., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujitake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matshiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
TITLE	4 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multiplexillary sequencer
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE	20530913
PUBMED	11076861
REFERENCE	4
AUTHORS	The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE	Functional annotation of a full-length mouse cDNA collection
JOURNAL	Nature 409, 685-690 (2001)
REFERENCE	5
AUTHORS	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL	Nature 420, 563-573 (2002)
MEDLINE	6 (bases 1 to 3012)
PUBMED	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hitomoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imocani, K., Iehii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Komno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murate, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ono, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, K., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shihagawa, A., Shikata, T., Sobabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yaeunishi, A., Muramatsu, M. and Hayashizaki, Y.
TITLE	Direct Submission
JOURNAL	Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physiological and Chemical Research (RIKEN), Laboratory for Genome Exploration and Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, fax: 81-45-503-9216)
COMMENT	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues.
Please visit our web site for further details.
URL: <http://genome.gsc.riken.go.jp/>
URL: <http://fantom.gsc.riken.go.jp/>
Location/Qualifiers

FEATURES.

source

1. .3012
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="PANTOM.DB:7330435N05"
/db_xref="MGI:2396891"
/db_xref="taxon:10090"
/clone="7330435N05"
/sex="male"
/tissue_type="adrenal gland"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
1. .3012
/note="B-cell translocation gene 3 (MGD)|MGI:109532,
GB|NM_009770, evidence: BLASTN, 99%, match=438)"

ORIGIN

Alignment Scores:

Pred. No.: 5.11e+04 Length: 3012

Score: 8.00 Matches: 8

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 1.20% Indels: 0

DB: 11 Gaps: 0

US-10-029-345A-109 (1-665) x AK078652 (1-3012)

QY 575 GYSerAlaSerTYrSerAlaTYr 582
|||||
|||||

DB 2395 GGCTCCGATCATACTCTGCTTAC 2372

RESULT 41
AK030955

LOCUS AK030955 3359 bp mRNA linear HTC 18-SEP-2003

DEFINITION Mus musculus adult male thymus cDNA, RIKEN full-length enriched library, clone:5830469C09 Product:similar to RAB3-GAP REGULATORY DOMAIN [Homo sapiens], full insert sequence.

ACCESSION AK030955 AK030955 GI:26082074

VERSION AK030955

KEYWORDS HTC, CAP trapper.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Weich. Enzymol. 303, 19-44 (1999)

AUTHORS Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

TITLE Normalization and substructure of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL Genome Res. 10 (10), 1617-1630 (2000)

MEDLINE 20499374

PUBMED 11042159

REFERENCE 3

AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Kono, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsuno, H., Sakauchi, S., Ikegami, T., Kashiwagi, K., Fujikawa, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Matsuki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.

TITLE RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer

COMMENT

URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: <http://genome.gsc.riken.go.jp/>
 URL: <http://fantom.gsc.riken.go.jp/>.

FEATURES

source

CDS

```

1..3335
   /organism="Mus musculus"
   /mol_type="mRNA"
   /strain="C57BL/6J"
   /db_xref="PANTOM DB:5930426E16"
   /db_xref="WGI:2354389"
   /db_xref="taxon:10090"
   /clone="5930426E16"
   /issue_type="forelimb"
   /clone_lib="RIKEN full-length enriched mouse cDNA library"
   /dev_stage="13 days embryo"
142..3147
   /note="unmated protein product; putative
   similar to RAB3-GAP REGULATORY DOMAIN [Homo sapiens]
   (SPTR|075872, evidence: FASTY, 87%id, 76.7%length,
   match=3138)"
   /codon_start=1
   /protein_id="BAC27296.1"
   /db_xref="GI:26327105"
   /translation="MEPATPLAVRFGILPDSRRHGESICLSPCNTLAATVDDEGRVILL
   DVARGIAIRWKGKRDQAVGMIVIEDLHRRVVEKGFSPFGNTQGPSRAQFLVYA
   PRGIEVNWVQGPGRVAFNVGKCHLLPGYKIMGNVNTQSOMPTQOICLVDP
   VSAQVAVNPFLIALSDKSKERAKDLHLYKTKALLRAKSPRSDFEFEIKELIDI
   KYPAIKQALLESIALSDPSLFCIRNVTOTIMTLTKNOLSEVDEBLTOPCKSKATL
   HLYESVQALNTLIDHSDTPSPNDLAVLRLDKELIKLRALKERIKQNTATVRS
   EDADRVLPVTPLELYEYEDALSIKVGEEBCVALGSFPWCKLHGSESTEMCHSL
   ESAGLSFQDLSTLSTLWSKEDKILDKPQSVCLATMTLSLKNMVAIDETWDSQV
   SPWQQRMAACIQSENSGALLSAHVAGSVAAQMSGADKFSQWVADADALATDS
   WEALISDTEYMKLLRLQLEDCILITLHLSKSPPAKAPLSQSEPLPVSVKLIKLG
   GKGIADSVAKWIFKODLSPELLKCANKEKRDVNPEDRDLHLAVRPPSCLIDY
   LHAHCMEVYVONKNDPEARPLVRSIEHLKOTLNHYONGALMMMTFLYKRSAA
   TYLMDKVKSPKRLCRDVGMSDTLITFSLSCBELTSLSEADISREVOVPVLT
   EDWLVSEGPISIVELALQKPTHPVLEHSHVLSLITSLASMRFLSKSVKPLADPSK
   GKNAFKDILTSIQILPSGEMDNFISVRQOFLKVVSAVQHQSHKDKPSAEANTH
   WKDLNPGIAVLDAHLQVSDVIRRHVYGEIYSGADLGEALFOYODKREVLASOL
   LVITCGRLAHLPHOTKEGMELIARLPITLCTWIKAMNPOLQNTGVPDIATATLVH
   KMWELPERKGGQSLALHLIDVNEAMATL"

```

ORIGIN

Alignment Scores:

Pred. No.: 5.69e+04 Length: 3335
 Score: 8.00 Matches: 8
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.20% Indels: 0
 DB: 11 Gaps: 0

US-10-029-345A-109 (1-665) x AK031191 (1-3335)

QY 309 SerTysLeuTylLeuLeuHsiLeu 316
 |||||
 952 AGTAATTAACTACTACATCT 975

RESULT 43

AK029006/c

LOCUS

DEFINITION

AK029006 3610 bp mRNA linear HTC 18-SEP-2003
 Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched
 library, clone:432460N13 product:naa (Hsp40) homolog, subfamily
 C, member 5, full insert sequence.

ACCESSION

AK029006
 AK029006.1 GI:26324963

KEYWORDS

HTC; CAP trapper.

SOURCE

ORGANISM

Mus musculus (house mouse)

Mus

musculus

REFERENCE

AUTHORS

1 Carninci, P. and Hayashizaki, Y.

TITLE

High-efficiency full-length cDNA cloning

JOURNAL

Meth. Enzymol. 303, 19-44 (1999)

MEDLINE

99279253

PUBMED

10349636

REFERENCE

AUTHORS

2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

TITLE

Normalization and subtraction of cap-trapper-selected cDNAs to
 prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL

Genome Res. 10 (10), 1617-1630 (2000)

MEDLINE

20499374

PUBMED

11042159

REFERENCE

AUTHORS

3 Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 Kono, H., Akiyama, J., Nishi, K., Kitsuai, T., Tashiro, H., Itoh, M.,
 Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,
 Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kasai, K.,
 Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,
 Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,
 Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.

TITLE

RIKEN integrated sequence analysis (RISA) system-384-format
 sequencing pipeline with 384 multicapillary sequencer

JOURNAL

Genome Res. 10 (11), 1757-1771 (2000)

MEDLINE

20530913

PUBMED

11076861

REFERENCE

AUTHORS

4 The RIKEN Genome Exploration Research Group Phase II Team and the
 PANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)

TITLE

5 The PANTOM Consortium and the RIKEN Genome Exploration Research
 Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)

JOURNAL

6 (bases 1 to 3610)

REFERENCE

6 (bases 1 to 3610)

AUTHORS

Adachi, J., Aizawa, K., Akimura, T., Arikawa, T., Bono, H., Carninci, P.,
 Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
 Hayashida, K., Hayashi, K., Hiramoto, K., Hiraoka, T., Hirozane, T.,
 Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
 Katoh, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,
 Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku, K., Takahira, S.,
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
 Muramatsu, M. and Hayashizaki, Y.

TITLE

Direct Submission

JOURNAL

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehito-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
 URL: <http://genome.gsc.riken.go.jp/>, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)

COMMENT

cDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.

Please visit our web site for further details.

URL: <http://genome.gsc.riken.go.jp/>
 URL: <http://fantom.gsc.riken.go.jp/>.

FEATURES

source

1..3610

```
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:4732480N13"
/db_xref="MGI:2391035"
/db_xref="taxon:10090"
/clone="4732480N13"
/tissue_type="skin"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="10 days neonate"
163..759
/notice="unnamed protein product; DnaJ (Hsp40) homolog,
subfamily C, member 5 (MGD|MGI:892995, GI|NW_016775,
evidence: BLASTN, 99%, match=597)
putative"
/codon_start=1
/protein_id="BAC26236.1"
/db_xref="GI:26324964"
/translation="MADQRQSLSTGSESLHYVLGLDKNATSDDIKIKYRLKLYHP
DKNPDEADKFEKINNAAILTDATKNRIYDKSGISGLYVGEQGEENVTVPLIS
SWAKLFLVYCGILTCYCCGCCGCCCKCKRPRGSEETERYVSPEDLEAQLQ
SDERATDPIYIQPASATETTTQLTIDSHPSYHTDGFN"
```

ORIGIN

Alignment Scores:

Pred. No.:	6.19e+04	Length:	3610
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.20%	Indels:	0
DB:	11	Gaps:	0

US-10-029-345A-109 (1-665) x AK029006 (1-3610)

QY 95 Val1aSer1euSer1aspCys 102
|||||
662 GTAGCCTCCTTCATCAGACTGC 639

RESULT 44
AK046988

LOCUS AK046988

DEFINITION Mus musculus 10 days neonate cerebellum cDNA, RIKEN full-length enriched library, clone:B930009J12 product:unclassified, full insert sequence.

ACCESSION AK046988

VERSION AK046988.1 GI:26091852

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1
Carninci,P. and Hayashizaki,Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)

JOURNAL MEDLINE 99279253

PUBMED 10349636

REFERENCE 2
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL MEDLINE 20499374

PUBMED 11042159

REFERENCE 3
Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kikunaka,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishino,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanishi,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsumura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kita,A. and Hayashizaki,Y.

TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer

JOURNAL MEDLINE 20530913

PUBMED 11076861

REFERENCE 4
The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

JOURNAL MEDLINE 11076861

PUBMED 11076861

REFERENCE 5
The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

JOURNAL MEDLINE 11076861

PUBMED 11076861

REFERENCE 6
(bases 1 to 3617)
Adachi,J., Aizawa,K., Akimura,T., Arikawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,M., Hayashida,K., Hayatsu,N., Hiramoto,K., Hiroaka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,T., Kasukawa,T., Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohnato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Kahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
Direct Submision
Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Teurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)

JOURNAL MEDLINE 11076861

PUBMED 11076861

COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL:http://genome.gsc.riken.go.jp/
URL:http://fantom.gsc.riken.go.jp/
URL:http://fantom.qualifiers

FEATURES
source
1..3617
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:B930009J12"
/db_xref="MGI:2412520"
/db_xref="taxon:10090"
/clone="B930009J12"
/tissue_type="cerebellum"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="10 days neonate"
1..3617
/note="unclassified"

misc_feature
/note="unclassified"

ORIGIN

Alignment Scores:

Pred. No.:	6.21e+04	Length:	3617
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.20%	Indels:	0
DB:	11	Gaps:	0

US-10-029-345A-109 (1-665) x AK046988 (1-3617)

QY 163 LeuProAsnLeuTyrIleuGlyCys 170
|||||

DB 3401 TTACCAACTGTATCTAGATGT 3424

RESULT 45
AK032373/c

LOCUS AK032373 3626 bp mRNA linear HTC 18-SEP-2003

DEFINITION Mus musculus adult male olfactory brain cDNA, RIKEN full-length enriched library, clone:6430529B06 product:DnaJ (Hsp40) homolog, subfamily C, member 5, full insert sequence.

ACCESSION AK032373

VERSION AK032373.1 GI:26328200

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS Carninci, P. and Hayashizaki, Y.

TITLE High-efficiency full-length cDNA cloning

JOURNAL Meth. Enzymol. 303, 19-44 (1999)

MEDLINE 99279253

PUBMED 10349636

REFERENCE 2

AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komano, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes

JOURNAL Genome Res. 10 (10), 1617-1630 (2000)

MEDLINE 20499374

PUBMED 11042159

REFERENCE 3

AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagao, S., Sasaki, N., Carninci, P., Komano, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Smil, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watanabe, M., Okazaki, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsushita, S., Kawai, J., Yoneda, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.

TITLE RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer

JOURNAL Genome Res. 10 (11), 1757-1771 (2000)

MEDLINE 20530913

PUBMED 11076861

REFERENCE 4

AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.

TITLE Functional annotation of a full-length mouse cDNA collection

JOURNAL Nature 409, 685-690 (2001)

MEDLINE 11076861

PUBMED 11076861

REFERENCE 5

AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.

TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL Nature 420, 563-573 (2002)

MEDLINE 11076861

PUBMED 11076861

REFERENCE 6

AUTHORS Adachi, J., Aizawa, K., Akimura, T., Arikawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Katsukawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Komano, H., Konda, M., Koyama, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, K., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, Y., Tanaka, T., Tomaru, A., Toyu, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.

TITLE Direct Submission

JOURNAL Submitted (16-Jul-2001) Yoshinide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsukuba-shi, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,

COMMENT

Fax:81-45-503-9216)

CDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site for further details.

URL: http://genome.gsc.riken.go.jp/

URL: http://fantom.gsc.riken.go.jp/

Location/Qualifiers

1.3626

source

1.3626

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="PANTOM DB:6430529B06"

/db_xref="MG1:235854"

/db_xref="taxon:10090"

/clone="6430529B06"

/sex="male"

/tissue_type="olfactory brain"

/clone_lib="RIKEN full-length enriched mouse cDNA library"

/dev_stage="adult"

157_-753

/note="unnamed protein product; DnaJ (Hsp40) homolog, subfamily C, member 5 (MG1:235854, GB|NM_016775, evidence: BLASTN, 99%, match=597)

putative"

/codon_start=1

/protein_id="PAC27841.1"

/db_xref="GI:26328201"

/translation="MAPQQRQSRSTSGSLVYVIGLDKRNATSDIKSKYKALKATPH DKRDNPEADKAEIINNAHILTDATKRNITDYKSLGVVAGQFEENVNYFVLS SWAKALFVCCGLTCCYCCCLCCCFCCGCKPKAPGSGETFEFVSPEDLEAQLQ SDEREADTPVLPASATETQTLTADSHPSHYTDGFN"

ORIGIN

Alignment Scores:

Pred. No.: 6.22e+04 Length: 3626

Score: 8.00 Matches: 8

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 1.20% Indels: 0

DB: 11 Gaps: 0

US-10-029-345A-109 (1-665) x AK032373 (1-3626)

Oy 95 ValAlaSerLeuSerSerAspCys 102

DB 656 GTAGCTCTCCCTTCATCAGACTGC 633

RESULT 46

LOCUS AK083783 3701 bp mRNA linear HTC 20-SEP-2003

DEFINITION Mus musculus 12 days embryo spinal ganglion cDNA, RIKEN full-length enriched library, clone:DL30008F08 product:unclassified, full insert sequence.

ACCESSION AK083783

VERSION AK083783.1 GI:26101489

KEYWORDS HTC; CAP trapper.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS Carninci, P. and Hayashizaki, Y.

TITLE High-efficiency full-length cDNA cloning

JOURNAL Meth. Enzymol. 303, 19-44 (1999)

MEDLINE 99279253

PUBMED 10349636

REFERENCE 2

AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komano, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.

TITLE Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

3
11042159

Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuami, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiiwaka, S., Inoue, K., Togawa, K., Izawa, M., Ohara, E., Watabiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
RIKEN Integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

4
11076861

The RIKEN Genome Exploration Research Group Phase II Team and the
PANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

5
10349636

The PANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

CDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.

FEATURES
source

1. 3701
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM_DB:D130008F08"
/db_xref="WGI:2419339"
/db_xref="taxon:10090"
/clone="D130008F08"
/issue_type="spinal ganglion"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="12 days embryo"
1. 3701
/note="unclassified"

ORIGIN

Alignment Scores:

Prod. No.: 6.36e+04 Length: 3701

Score: 8.00 Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 1.20% Indels: 0

DB: 11 Gaps: 0

US-10-029-345a-109 (1-665) x AK083783 (1-3701)

QY 359 SerValProSerValProSerVal 366

Db 2227 TCTGTGCCCTCTGTGCCCTCTGTG 2250

RESULT 47

AK084916

LOCUS

DEFINITION

AK084916

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

AK084916 3722 bp mRNA linear HTC 20-SEP-2003
Mus musculus 13 days embryo lung cDNA, RIKEN full-length enriched
library, clone:D430011E20 product:hypothetical protein, full insert
sequence.
AK084916.1 GI:26351342
AK084916.1 HTG: CAP trapper.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)

2
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)

3
Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
Konno, H., Akiyama, J., Nishi, K., Kitsuami, T., Tashiro, H., Itoh, M.,
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,
Fujiiwaka, S., Inoue, K., Togawa, K., Izawa, M., Ohara, E., Watabiki, M.,
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J.,
Okazaki, Y., Muramatsu, M., Inoue, Y., Kita, A. and Hayashizaki, Y.
RIKEN Integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

4
The RIKEN Genome Exploration Research Group Phase II Team and the
FANTOM Consortium.
Functional annotation of a full-length mouse cDNA collection
Nature 409, 685-690 (2001)

5
The PANTOM Consortium and the RIKEN Genome Exploration Research
Group Phase I & II Team.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

6
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
Fax: 81-45-503-9216)

CDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.

TITLE
JOURNAL

COMMENT

FEATURES
source

Okazaki, Y., Satoh, R., Satoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akashira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y.
Direct Submission
Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration and Chemical Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Tsurumi-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gsc.riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: http://genome.gsc.riken.go.jp/
URL: http://fantom.gsc.riken.go.jp/.
Location/Qualifiers

CDS

ORIGIN

Alignment Scores:

Pred. No.: 6.4e+04 Length: 3722
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 11 Gaps: 0

US-10-029-345A-109 (1-665) x AK084916 (1-3722)

OY 357 ProAlaSerValProSerValPro 364

DB 1207 CCAGCGCTGTGTAAGCTTCTGTACT 1230

RESULT 48

BC049163 3722 bp mRNA linear HTC 17-DEC-2003
DEFINITION Mus musculus cDNA clone IMAGE:6466814, containing frame-shift errors.
ACCESSION BC049163
VERSION BC049163.1 GI:29165699
KEYWORDS HTC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE
AUTHORS

Strusberg, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,

TITLE

JOURNAL

PUBMED

AUTHORS

JOURNAL

REMARK
COMMENT

Altechul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhattacharya, N.K., Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, P., Diachenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L., Stachleiron, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L., Scheetz, T.E., Brownstein, M.J., Ueda, T.B., Toshiyuki, S., Cantini, P., Prange, C., Raha, S.S., Loguercio, N.A., Peters, G.J., Abramson, R.D., Mullaly, S.J., Bosak, S.A., McMan, P.J., McKernan, K.J., Malek, J.A., Gunnarsson, P.H., Richards, S., Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Huliyil, S.W., Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A., Fahy, J., Helton, E., Kettelman, M., Madan, A., Rodriguez, S., Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y., Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D., Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M., Butlerfield, Y.S., Kitzwinski, M.I., Skalska, U., Smalins, D.E., Schercher, A., Schein, J.E., Jones, S.J., Jones, S.J., Marra, M.A. Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
12477932
2 (bases 1 to 3722)
Strusberg, R.
Direct Submission
Submitted (21-MAR-2003) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
http://www.systemsbio.org
Contact: amadan@systemsbio.org
Anup Madan, Jessica Fahy, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodriguez, Amy Sanchez and Michelle Whiting
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
Series: IRAC Plate: 105 Row: 1 Column: 15
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA GI: 27369777
This clone has the following problem: frame shifted.
Location/Qualifiers

FEATURES

source

ORIGIN

Alignment Scores:

Pred. No.: 6.4e+04 Length: 3722
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 11 Gaps: 0

US-10-029-345A-109 (1-665) x BC049163 (1-3722)

OY 410 SerAlaSerValAlaAlaSerLeu 417

DB 2435 AGTGCATGATGGAGCAGCTT 2458

RESULT 49

AK035810

LOCUS AK035810 3727 bp mRNA linear HTC 19-SEP-2003
DEFINITION Mus musculus 16 days neonate cerebellum cDNA, RIKEN full-length enriched library, clone:9630006820 product:POLY(A) POLYMERASE GAMMA (EC 2.7.7.19) (NEO-POLY(A) POLYMERASE) homolog [Homo sapiens], full insert sequence.
ACCESSION AK035810
VERSION AK035810.1 GI:26330927
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Komno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159
REFERENCE 3
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagasaka, S., Sasaki, N., Carninci, P., Komno, H., Akiyama, J., Nishi, K., Kitanishi, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, S., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Yamada, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, O., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
TITLE RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE 20530913
PUBMED 11076861
REFERENCE 4
AUTHORS The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
TITLE Functional annotation of a full-length mouse cDNA collection
JOURNAL Nature 409, 685-690 (2001)
PUBMED 11076861
REFERENCE 5
AUTHORS The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
TITLE Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
JOURNAL Nature 420, 563-573 (2002)
PUBMED 11076861
REFERENCE 6
AUTHORS Aichi, J., Aizawa, K., Akimura, T., Arai, K., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirazawa, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kankawa, T., Kato, H., Kawai, J., Kojima, Y., Kondo, S., Komno, H., Kouda, M., Koya, S., Kurinara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Ohgato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaki, A., Tanaka, Y., Tanaka, T., Tomaru, A., Toyota, T., Yasunishi, A., Muramatsu, M., and Hayashizaki, Y.
TITLE Direct Submission
JOURNAL Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp, URL: <http://genome.gsc.riken.go.jp/>, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
Please visit our web site for further details.
URL: <http://genome.gsc.riken.go.jp/>
URL: <http://fantom.gsc.riken.go.jp/>
Location/Qualifiers
FEATURES
source
1..3727
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:9630006820"
/db_xref="MGI:2400157"
/db_xref="taxon:10090"
/clone="9630006820"
/tissue_type="cerebellum"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="16 days neonate"
216..2435
/note="unnamed protein product; POLY(A) POLYMERASE GAMMA (EC 2.7.7.19) (NEO-POLY(A) POLYMERASE) homolog [Homo sapiens] (SPR|Q969N1, evidence: PASTY, 93.5%ID, 100%length, match=2217)
putative"
/codon_start=1
/protein_id="BAC29194.1"
/db_xref="GI:26330928"
/translation="MKEMSANMTLDSQROOKHYGTPISLACPKEDIDHYTKLIDA MKPFGVEEELNHLVVLGKMLNVLKMSIDISSEKULPSVAVTGGKIPFGSY RGVHTKGADIDLCVAPRPHVRSDFOSFEKIKQDGRNLRAVEDAPVVKFES DGIETLVAPRLATISNLDRLSDRLSIDIPTISGRCRTVDILHYPNKT FRLTIRAVLAKAKRGITYSNMGFLGVSMAHIVATCOLYPRMAASTLVKHFVLYS KWEKPVVLKQPEESNLNLPVWDVNSDRHMLPITPAIPQONSTYVNSTRT VVWEVKQGLAVYDELILQKSDMSKLEPTNFOKRYHIVLASASTENHLEWGL VSKIRVAGLNEERNEFITLAVNPQSPFGNKHANNVSMFGLIIFRYENAES VNIDLVDIOSFDYVROANNINMKQDKIBATHVKKKLHYPALIELOKKSL SVSRSGGLQKRSLSLSDSCSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSDSA ERYAVVVELPSVPPAQGLSTPYTAKVDPTAKVNSPAVCTIPYGVNVRITTP HNPVQGPLHNGISNTTKVTNTPKRSPLPTDGSKLIKIEKTRILESAPKESRAED RRRKPMDSIGESMPITPTIDARKKRLPSKELEPDSPPVANNIRVINKSIRLTNR"

ORIGIN
Alignment Scores:
Pred. No.: 6,41e+04 Length: 3727
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 11 Gaps: 0
US-10-029-345A-109 (1-665) x AK035810 (1-3727)
Qy 410 Ser1ase7Me1a1ase7leu 417
Db 2474 AGTGCATCGATGCGACCACTCTT 2497
RESULT 50
ACCESSION AK078517
VERSION AK078517.1 GI:26097911
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning

JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636
 REFERENCE

AUTHORS

TITLE

JOURNAL Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,
 MEDLINE Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
 PUBMED Normalization and subtraction of cap-trapper-selected cDNAs to
 20499374 prepare full-length cDNA libraries for rapid discovery of new genes
 11042159 Genome Res. 10 (10), 1617-1630 (2000)

REFERENCE

AUTHORS

TITLE

JOURNAL Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,
 MEDLINE Kono, H., Akiyama, J., Nishi, K., Kitsumi, T., Tashiro, H., Itoh, M.,
 PUBMED Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,
 11076861 Yamamoto, R., Matsunoto, H., Sakaguchi, S., Ikegami, T., Kashwagi, K.,
 4 Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,
 Okazaki, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J.,
 RIKEN integrated sequence analysis (RISA) system--384-format
 sequencing pipeline with 384 multicapillary sequencer
 Genome Res. 10 (11), 1757-1771 (2000)

REFERENCE

AUTHORS

TITLE

JOURNAL The RIKEN Genome Exploration Research Group Phase II Team and the
 MEDLINE FANTOM Consortium.
 PUBMED Functional annotation of a full-length mouse cDNA collection
 5 Nature 409, 685-690 (2001)

REFERENCE

AUTHORS

TITLE

JOURNAL The FANTOM Consortium and the RIKEN Genome Exploration Research
 MEDLINE Group Phase I & II Team.
 PUBMED Analysis of the mouse transcriptome based on functional annotation
 6 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)

REFERENCE

AUTHORS

TITLE

JOURNAL Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
 MEDLINE Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
 PUBMED Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
 11076861 Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
 Kato, H., Kawai, J., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,
 Koyu, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
 Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
 Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
 Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
 Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
 Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
 Muramatsu, M., and Hayashizaki, Y.
 Direct Submission
 Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
 Physical and Chemical Research (RIKEN), Laboratory for Genome
 Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
 RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
 Kanagawa 230-0045, Japan (E-mail: genome-res@gs.riken.go.jp,
 URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
 Fax: 81-45-503-9216)

COMMENT

CDNA library was prepared and sequenced in Mouse Genome
 Encyclopedia Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in RIKEN.
 Division of Experimental Animal Research in Riken contributed to
 prepare mouse tissues.
 Please visit our web site for further details.
 URL: http://genome.gsc.riken.go.jp/
 URL: http://fantom.gsc.riken.go.jp/
 Location/Qualifiers

FEATURES

source

1. 3789
 /organism="Mus musculus"
 /mol_type="cDNA"
 /strain="CS7BL/6J"
 /db_xref="FANTOM_DB:682042BL09"
 /db_xref="MGI:2396740"
 /db_xref="taxon:10090"
 /clone="682042BL09"

misc_feature
 1. 3789
 /note="weakly similar to KIAA1614 PROTEIN (FRAGMENT) (Homo
 sapiens) (SPT|QSHCF8, evidence: FASTV, 57.3%ID,
 68.7%length, match=2504)"

ORIGIN

Alignment Scores:

Pred. No.:	6 52e+04	Length:	3789
Score:	8.00	Matches:	8
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.20%	Indels:	0
DB:	11	Gaps:	0

US-10-029-345A-109 (1-665) x AK078517 (1-3789)

QY 544 LeuAlaProGlnThrSerThrPro 551

DB 2171 CTGGCTCCACAGACGACGACCC 2194

Search completed: June 22, 2004, 10:09:34
 Job time : 437 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 22, 2004, 03:17:17 ; Search time 664 Seconds
(without alignments)
4254.596 Million cell updates/sec

Title: US-10-029-345a-109

Perfect score: 665

Sequence: 1 MAHEMIGCTIVTERLVALL.....LKVSGSQSFSGSMETIEVS 665

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 212409041 residues

Word size: 1

Total number of hits satisfying chosen parameters: 772722

Minimum DB seq length: 1419

Maximum DB seq length: 200000000

Post-processing: Listing first 75 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=xlh
-Q/cg2.1/USPTO.spc01/US10023345/runat.21062004.164401.29745/app.query.fasta.1.839
-DB=N.Geneseq.29Jan04 -QFMT=fastap -SUFFIX=oligo.rng -MINMATCH=0.1 -LCOOPT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi
-LIST=75 -DOCALLIGN=200 -THR SCORE=quality -THR MIN=1 ALIGN=50 -MODE=LOCAL
-OUTFMT=pro -NORM=ext -HEA=SIZE=500 -MINLEN=1419 -MAXLEN=200000000
-USER=US10023345.@CGN_1_1.470.@runat.21062004.164401.29745 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -NARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

N.Geneseq.29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	665	100.0	5450	6	ACCG0559	ACCG0559 Polynucle
2	665	100.0	5450	6	ACCG0572	ACCG0572 Polynucle
3	472	71.0	2732	7	ADA53105	ADA53105 Human cod
4	472	71.0	2732	4	AAD09492	AAD09492 Human SGP
5	472	71.0	2966	4	AAH99685	AAH99685 Human pro
6	472	71.0	3059	4	AAH99685	AAH99685 Human pro
7	472	71.0	3104	6	ABN59704	ABN59704 Novel hum
8	472	71.0	3332	6	ABK48378	ABK48378 cDNA enco

9	472	71.0	3496	6	ABK47596	ABK47596 cDNA enco
10	472	71.0	3544	5	AA514639	AA514639 Human cDN
11	472	71.0	3544	6	ABK49402	ABK49402 cDNA enco
12	472	71.0	3766	6	ABK14474	ABK14474 Human pro
13	472	71.0	4790	6	ABN83966	ABN83966 Human gen
14	472	71.0	5145	5	ABV20833	ABV20833 Human pro
15	472	71.0	5145	5	ABV21080	ABV21080 Human pro
16	472	71.0	5145	5	ABV26680	ABV26680 Human pro
17	472	71.0	5145	5	ABV20978	ABV20978 Human pro
18	472	71.0	5145	5	ABV21092	ABV21092 Human pro
19	472	71.0	5145	5	ABV21312	ABV21312 Human pro
20	472	71.0	5145	5	ABV21316	ABV21316 Human pro
21	472	71.0	5145	5	ABV26886	ABV26886 Human pro
22	472	71.0	5145	5	ABV27131	ABV27131 Human pro
23	472	71.0	5145	5	ABV26923	ABV26923 Human pro
24	472	71.0	5145	5	ABV27135	ABV27135 Human pro
25	472	71.0	5145	5	ABV28657	ABV28657 Human pro
26	472	71.0	5145	5	ABV22827	ABV22827 Human pro
27	472	71.0	5145	5	ABV26934	ABV26934 Human pro
28	472	71.0	5145	5	ABV26934	ABV26934 Human pro
29	472	71.0	5145	5	ABV26934	ABV26934 Human pro
30	472	71.0	5145	5	ABV26934	ABV26934 Human pro
31	472	71.0	5145	5	ABV26934	ABV26934 Human pro
32	472	71.0	5145	5	ABV26934	ABV26934 Human pro
33	472	71.0	5145	5	ABV26934	ABV26934 Human pro
34	472	71.0	5145	5	ABV26934	ABV26934 Human pro
35	472	71.0	5145	5	ABV26934	ABV26934 Human pro
36	472	71.0	5145	5	ABV26934	ABV26934 Human pro
37	472	71.0	5145	5	ABV26934	ABV26934 Human pro
38	472	71.0	5145	5	ABV26934	ABV26934 Human pro
39	472	71.0	5145	5	ABV26934	ABV26934 Human pro
40	472	71.0	5145	5	ABV26934	ABV26934 Human pro
41	472	71.0	5145	5	ABV26934	ABV26934 Human pro
42	472	71.0	5145	5	ABV26934	ABV26934 Human pro
43	472	71.0	5145	5	ABV26934	ABV26934 Human pro
44	472	71.0	5145	5	ABV26934	ABV26934 Human pro
45	472	71.0	5145	5	ABV26934	ABV26934 Human pro
46	472	71.0	5145	5	ABV26934	ABV26934 Human pro
47	472	71.0	5145	5	ABV26934	ABV26934 Human pro
48	472	71.0	5145	5	ABV26934	ABV26934 Human pro
49	472	71.0	5145	5	ABV26934	ABV26934 Human pro
50	472	71.0	5145	5	ABV26934	ABV26934 Human pro
51	472	71.0	5145	5	ABV26934	ABV26934 Human pro
52	472	71.0	5145	5	ABV26934	ABV26934 Human pro
53	472	71.0	5145	5	ABV26934	ABV26934 Human pro
54	472	71.0	5145	5	ABV26934	ABV26934 Human pro
55	472	71.0	5145	5	ABV26934	ABV26934 Human pro
56	472	71.0	5145	5	ABV26934	ABV26934 Human pro
57	472	71.0	5145	5	ABV26934	ABV26934 Human pro
58	472	71.0	5145	5	ABV26934	ABV26934 Human pro
59	472	71.0	5145	5	ABV26934	ABV26934 Human pro
60	472	71.0	5145	5	ABV26934	ABV26934 Human pro
61	472	71.0	5145	5	ABV26934	ABV26934 Human pro
62	472	71.0	5145	5	ABV26934	ABV26934 Human pro
63	472	71.0	5145	5	ABV26934	ABV26934 Human pro
64	472	71.0	5145	5	ABV26934	ABV26934 Human pro
65	472	71.0	5145	5	ABV26934	ABV26934 Human pro
66	472	71.0	5145	5	ABV26934	ABV26934 Human pro
67	472	71.0	5145	5	ABV26934	ABV26934 Human pro
68	472	71.0	5145	5	ABV26934	ABV26934 Human pro
69	472	71.0	5145	5	ABV26934	ABV26934 Human pro
70	472	71.0	5145	5	ABV26934	ABV26934 Human pro
71	472	71.0	5145	5	ABV26934	ABV26934 Human pro
72	472	71.0	5145	5	ABV26934	ABV26934 Human pro
73	472	71.0	5145	5	ABV26934	ABV26934 Human pro
74	472	71.0	5145	5	ABV26934	ABV26934 Human pro
75	472	71.0	5145	5	ABV26934	ABV26934 Human pro

RESULT 1

ACCG0559

ALIGNMENTS

XX	AD	ACC60559 standard; cDNA, 5450 BP.	
XX	AC	ACC60559;	
XX	DT	19-JUN-2003 (first entry)	
XX	DE	Polynucleotide relating to the invention SEQ ID NO: 108.	
XX	KM	Gene; ss; antiproliferative; hepatotropic; nephrotropic; antiarthritis;	
XX	KM	antipsoiatric; cardiant; cytostatic; gene therapy; liver disease;	
XX	KM	immunological disorder; arthritis; psoriasis; congenital heart defect;	
XX	KM	congestive heart failure; cancer; human; mouse; yeast; rat; fruitfly.	
XX	OS	Homo sapiens.	
XX	PN	WO200257460-A2.	
XX	PD	25-JUL-2002.	
XX	PF	20-DEC-2001; 2001WO-US050459.	
XX	PR	20-DEC-2000; 2000US-0256868P.	
XX	PR	30-MAR-2001; 2001US-0280186P.	
XX	PR	01-MAY-2001; 2001US-0287735P.	
XX	PR	05-JUN-2001; 2001US-0295848P.	
XX	PR	25-JUN-2001; 2001US-0300465P.	
XX	PA	(BRIM) BRISTOL-MYERS SQUIBB CO.	
XX	PI	Jackson DG, Feder J, Nelson T, Mintier G, Ramanathan C, Lee L;	
XX	PI	Siemers N, Bol D, Schieven G, Finger J, Todderud CG, Bassolino D;	
XX	PI	Krytek S, Mcatee P, Suchard S, Banas D;	
XX	DR	WPI: 2002-599721/64.	
XX	DR	P-PSDB; ABR52381.	
XX	PT	Novel polynucleotides encoding human phosphatase polypeptides useful in	
XX	PT	the prevention or treatment of e.g. proliferative and cardiovascular	
XX	PT	disorders.	
XX	PS	Claim 1; Fig 13; 80pp; English.	
XX	CC	The invention relates to a novel isolated nucleic acid comprising a	
XX	CC	polynucleotide having a nucleotide sequence selected from 40	
XX	CC	polynucleotides fully defined in the specification. The polynucleotide of	
XX	CC	the invention has antiproliferative, hepatotropic, nephrotropic, the	
XX	CC	antiarthritic, antipsoiatric, cardiant, and cytostatic activity. The	
XX	CC	polynucleotide may have a use in gene therapy. A polynucleotide or	
XX	CC	polypeptide of the invention is useful for preventing, treating or	
XX	CC	ameliorating a medical condition, e.g. a proliferative disorder. They are	
XX	CC	also useful for treating e.g. liver disease, renal failure, immunological	
XX	CC	disorders including arthritis and psoriasis, cardiovascular disorders	
XX	CC	such as congenital heart defects and congestive heart failure, and	
XX	CC	cancer. A method of the invention is useful for diagnosing a pathological	
XX	CC	condition or susceptibility to a condition in a subject. The present	
XX	CC	sequence is used in the exemplification of the invention	
XX	XX	Sequence 5450 BP; 1471 A; 1211 C; 1337 G; 1431 T; 0 U; 0 Other;	
XX	XX	Alignment Scores:	
XX	XX	Pred. No.: 0 Length: 5450	
XX	XX	Score: 665.00 Matches: 665	
XX	XX	Percent Similarity: 100.00% Conservative: 0	
XX	XX	Best Local Similarity: 100.00% Mismatches: 0	
XX	XX	Query Match: 100.00% Indels: 0	
XX	XX	DB: 6 Gaps: 0	
XX	XX	US-10-029-345A-109 (1-665) x ACC60559 (1-5450)	
XX	XX	0Y 1 MetAlAHISGLUmetctleGlyThrGlnIleValThrIuArgLeuValAlaLeuLeuGlu 20	
XX	XX	538 ATGGCCCATGAGATGATTTGGAACTCAATTGTATCTGAGAGTGGCGCTTGCTGGAA 597	

QY	21	SerGlyYThrGluLysValLeuLeuIleAapSerArgProPheValGluIrrhAntSer	40
Db	598	AGTGAACGAAAMAATGTGCTAAATTGATAGCGGCGCATTTGTGGAAATACAAATACATCC	657
QY	41	HisIleLeuGluValIleAsnIleAenCySerLysLeuMetLysArgArgLeuGlnGln	60
Db	658	CACATTTTGGAAAGCATTTATATATCAACTCTCCAACTTATATGAGGAAAGTTGGCAAG	717
QY	61	AapLysValLeuIleThrGluLeuIleGlnHisSerAlaLysHisLysValAspIleAap	80
Db	718	GACAAAGCTTATATTAACAGACTCACTCCAGCATTCAGCGAAACATPAGTTGACATTTGAT	777
QY	81	CysSerGlnLysValValValIrrhArgIleSerSerGlnAspValAlaSerLeuSerSer	100
Db	778	TGCACTCAAGAGGTGTGATTTACATCAAACTCCCAAGATGTGCTCTCTCTCTTAA	837
QY	101	AapCySphenLeuThrValLeuLeuGluLysLeuGluLysSerPheAsnSerValHisLeu	120
Db	838	GACTGTTTTCTACAGTACTTCTGGGTAACTGGAGAAGAGTTCACTCTTTCACCTG	897
QY	121	LeuAlaGlyGlyPheAlaGluPheSerArgCysPheProGlyLeuCySgluGlyLysSer	140
Db	898	CTTGCAGGGGGTGTCTGAGTTCCTCTGTGTTCCTCGGCGCTGTGGAAGAAATCC	957
QY	141	ThrLeuValProThrCysIleSerGlnProCysLeuProValAlaAsnIleGlyProThr	160
Db	958	ACTCTAGTCCCTCACTGCACTTTCTAGCCTTCTTACTCTGTGGCCAAACATTGGGCGAAC	1011
QY	161	ArgIleLeuProAsnLeuIrrhLeuGlyCysGlnArgAspValLeuAsnLysGluLeuIle	180
Db	1018	CGAAATCTCTCCAAATCTTATCTTGCGTGGCCAGCAGATGTCTCAACAAGAGCTGATTA	1077
QY	181	GlnGlnAsnGlyIleGlyIrrhValLeuAsnAlaSerIrrhCysProLysProAspPhe	200
Db	1078	CAGCGAAATGGGATTTGTTATGTGTAAATGCAAGCTATACCTGTCCAAAGCTGACATT	1137
QY	201	IleProGlnSerHisPheLeuArgValProValAsnAapSerPheCysGluLysIleLeu	220
Db	1138	ATCCCGAGTCTCAATTCCTCGGTGTGCTGTGAATGACACACTTTGTGAGAAATTTTG	1197
QY	221	ProTrpLeuAapLysSerValAspPheIleGluLysAlaLysAlaSerAsnGlyCysVal	240
Db	1198	CCGTGGTGGCAAAATCAGTAGATTTCATTAGAAAGCAAAAGCCTCCAAAGGATGTGCT	1257
QY	241	LeuValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrrIleMet	260
Db	1258	CTAGTGCACCTGTTAGCTGGGATCTCCGCTCCGCCACCAATCGCTATGCTTACATATG	1317
QY	261	LysArgMetAapMetSerLeuAspGluAlaTyrrArgPheValLysGluLysArgProThr	280
Db	1318	AAGAGGATGGACATGCTCTTTAGATGAAGCTTACAGATTTTGGAAAGAAAAAGACTTACT	1377
QY	281	IleSerProAsnPheAsnPheLeuGlyGluLeuLeuAspArgGluLysIleLysAsn	300
Db	1378	ATATCTCCAACTTCAATTTCTGGGCCAACTCTCGACTATGAGAAAGAAATTAAAGAAC	1437
QY	301	GlnThrGlyAlaSerGlyProLysSerLysLeuLysLeuLeuHisLeuGluLysProAsn	320
Db	1438	CAGACTGGAGCATCAGGGCCAAAGAGCAAACTCAAGGTGCTGCACCTGGAGAAACCAAT	1497
QY	321	GluProValProAlaValSerGluGlyGlyGluLysSerGluThrProLeuSerProPro	340
Db	1498	GAACCTGTCTCTGTGCTCAGAGGGTGGACAGAAAAAGCAGAGCCCTCACTGCTCACCC	1557
QY	341	CysAlaAspSerAlaTrnSerGluAlaAlaGlyGlnArgProValHisSerProAlaSerVal	360
Db	1558	TGTGCGCACTGTCTTACTCAGAGGACGACGAGCAAAAGCCCGTGCATCCCGCCAGCGTG	1617
QY	361	ProSerValProSerValGlnProSerLeuLeuLysAspSerProLeuValGlnAlaLeu	380
Db	1618	CCGACGTGGCCAGCGGTGACCGGTGCTGTGTGAGACAGCCGCGTGTGATACAGCGCTC	1677

QY	381	SerGlyLeuNH ₂ LeuSerAlaAspArgGluGluAspSerAsn ₁ LeuLYsArgSerPhe	400			
Db	1678	AGTGGCGCTGCACCTGTCCGACAGACGGCTGGAAAGACAAATTAAGCTCAACGCTTCTTC	1737			
QY	401	SerLeuAspIleLYsSerValSerTYrSerAlaSerMetAlaAlaSerLeuNH ₂ GlyPhe	420			
Db	1738	TCTCTGGATATCAATCAGTTTTCATATTACAGCCAGACATGGCAGCATCTTTCATGGGCTTC	1797			
QY	421	SerSerSerGluAspAlaLeuGluTYrTYrLYsProSerThrThrLeuAspGlyTYrAsn	440			
Db	1798	TCCCTCATCAGAAAGATGCTTTGGAAATACTACAAACCTTCCACTACTCTGGATGGAGCAAC	1857			
QY	441	LYsLeuCYsGlnPheSerProValGlnGluLeuSerGluGlnProGluTYrThrSerPro	460			
Db	1858	AAGCTATGCCAGTTCTCCCTGTTTCAGGAACATTCGGAGACTCCGAAACCAAGTCTCT	1917			
QY	461	AspLYsGluGluAlaSerIleProLYsLYsLeuGlnThrAlaArgProSerAspSerGln	480			
Db	1918	GATAAGAGAGAAAGCCAGCATCCCAAGAAAGCTGCAGACCGGCAAGCTTCAGACAGCCAG	1977			
QY	481	SerLYsArgLeuNH ₂ AsSerValArgThrSerSerSerGlyThrAlaGlnArgSerLeuLeu	500			
Db	1978	AGCAAGCCATTTGCAATTCGGCTCAGAAACCGACAGAGTGGCACCGCCGACAGAGCTCCCTTTTA	2037			
QY	501	SerProLeuNH ₂ AsArgSerGlySerValGluAspAsnTYrHisThrSerPheLeuPheGly	520			
Db	2038	TCTCCACATCGATCGAAGTGGAGCGCTGGAGACAATTACACACACAGCTTCCTTTTCGGC	2097			
QY	521	LeuSerThrSerSerGlnGlnHisLeuThrLYsSerAlaGlyLeuGlyLYsLeuGlyTYrPheHis	540			
Db	2098	CTTTCACCAACCCAGACAGACCTTACGAAAGTCTGCTGGCTTGGGCTTAAAGGCTGGCAC	2157			
QY	541	SerAspIleLeuAlaProGlnThrSerThrProSerLeuThrSerSerTYrTYrPheAla	560			
Db	2158	TCGGATATCTGGCCCCCAGACCTCTACCCCTTCCGACACAGAGCTGGTATTTTGCC	2217			
QY	561	ThrGluSerSerHisPheTYrSerSerAlaSerAlaIleTYrGlyLYsSerAlaSerTYrSer	580			
Db	2218	ACAGAGTCTCTACACTTCTACTCTGCCCTCACCCACATTCAGGAGGACAGCCAGATTACTCT	2277			
QY	581	AlaTYrSerCYsSerGlnLeuProThrCYsGlyAspGlnValTYrSerValArgArgArg	600			
Db	2278	GCTTACAGCTGCAGCCAGCTGCTGCCACTTGGCGAGACCAAGTCTATTCTGTCCGAGCGG	2337			
QY	601	GlnLYsProSerAspArgAlaAspSerArgArgSerTYrPheIleGluGluSerProPheGlu	620			
Db	2338	CAGAAAGCCAAAGTACAGAGCTGACTCGCGGGGAGCTGCATGAAGAAGGCCCTTTGAA	2397			
QY	621	LYsGlnPheLYsArgArgSerCYsGlnMetGluPheGlyGluSerIleMetSerGluAsn	640			
Db	2398	AAGCAGTTTAAACCGAGAAAGCTGCCAAATGTGAATTGGAGAGAGCATCATGTCCAGAAC	2457			
QY	641	ArgSerAspGluGluLeuGlnLYsValGlySerGlnSerSerPheSerGlySerMetGlu	660			
Db	2458	AGGTACCGGAGAAAGCTGGGAGAAAGTGCGAGCTGACTTCTTCGGGACAGCATGGAA	2517			
QY	661	IleIleGluValSer 665				
Db	2518	ATCATTTAGAGTCTCC 2532				
RESULT 2						
ACC60572						
ID	ACC60572 standard, cDNA, 5450 BP.					
NC	ACC60572;					
XX	19-JUN-2003 (first entry)					
DT						
XX						
DE	Polynucleotide relating to the invention SEQ ID NO: 147.					
XX						
Gene, ss; antiproliferative; hepatotropic; nephrotropic; antiarthritic;						
antipneumatic; cardiant; cyostatic; gene therapy; liver disease;						
proliferative disorder; renal failure; cardiovascular disorder;						

KW		immunological disorder; arthritis; psoriasis; congenital heart defect;
XV		congestive heart failure; cancer; human; mouse; yeast; rat; fruitfly.
XX		
OS	Homo sapiens.	
XX		
PN	WO200257460-A2.	
XX		
PD	25-JUL-2002.	
XX		
PF	20-DEC-2001; 2001WO-US050459.	
XX		
PR	20-DEC-2000; 2000US-0256866P.	
PR	30-MAR-2001; 2001US-0280186P.	
PR	01-MAY-2001; 2001US-0287735P.	
PR	05-JUN-2001; 2001US-0295848P.	
PR	25-JUN-2001; 2001US-0300465P.	
XX		
PA	(BRIM) BRISTOL-MYERS SQUIBB CO.	
PI	Jackson DG, Feder J, Nelson T, Mintier G, Ramanathan C, Lee L,	
PI	Siemers N, Bol D, Schieven G, Finger J, Todderud CG, Basolino D;	
PI	Krystek S, Meatee P, Suchard S, Banas D;	
XX		
DR	WPT; 2002-59972L/64.	
XX	P-PsDB; ABR52407.	
XX		
PT	Novel polynucleotides encoding human phosphatase polypeptides useful in	
PT	the prevention or treatment of e.g. proliferative and cardiovascular	
PT	disorders.	
XX		
PS	Example 7, Fig 19, 801pp; English	

```

CC The invention relates to a novel isolated nucleic acid comprising a
CC polynucleotide having a nucleotide sequence selected from 40
CC polynucleotides fully defined in the specification. The polynucleotide of
CC the invention has antiproliferative, hepatotropic, nephrotropic,
CC antitachytic, antipsoarctic, cardiant, and cytostatic activity. The
CC polynucleotide may have a use in gene therapy. A polynucleotide or
CC polypeptide of the invention is useful for preventing, treating or
CC ameliorating a medical condition, e.g. a proliferative disorder. They are
CC also useful for treating e.g. liver disease, renal failure, immunological
CC disorders including arthritis and psoriasis, cardiovascular disorders
CC such as congenital heart defects and congestive heart failure, and
CC cancer. A method of the invention is useful for diagnosing a pathological
CC condition or susceptibility to a condition in a subject. The present
CC sequence is used in the exemplification of the invention
XX
SQ Sequence 5450 BP, 1471 A, 1211 C, 1337 G, 1431 T, 0 U, 0 Other;

Alignment Scores:
Pred. No.: 0 Length: 5450
Score: 665.00 Matches: 665
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-029-345A-109 (1-665) x ACC60572 (1-5450)

QY 1 MetAlaHisgIuMetilegIyThrgInilevalThrgIuArgLeuValAlaleuLeuglu 20
Db 538 ATGGCCCATGAGATATTGGAACCTCAAAATTGTACTGAGAGCTTGCTGCTCGGAA 597
QY 21 SerGIyThrgIuArgValleuLeuLeuIleApsSerATcGpProPhaValgIuIyRanThrSer 40
Db 598 AGTGAACCGAANAAGTGTCTGCTTAATTGATTCACCGCCATTGTGGAATAACATACATCC 657
QY 41 HisIleLeugIuAlaIleAsnIleAncySerIySleuMetIyArghArgIeugIngin 60
Db 658 CACATTTTGGAAGCATTATATATCAACGTCTCCAAAGCTTATGAGAGGAAGTTGCAACAG 717
QY 61 AspyIyValleuIleThrgIuLeuIleGlnIleSerAlaIyGhisIyValAspIleAap 80

```

Db 718 GACAAAGTGTAAATTACAGAGCTCATCCAGCATTCAGCAAAACATAAGTTGACATTGAT 777
 QY 81 CYSSEGLILYVALVALVALYTRAPGINSERSEGLINAPVALALASERLEUSER 100
 Db 778 TCACAGTCAAGAGTGTAGTTTACATCAAAAGCTCCCAAGATGTTGCCCTCTCTCTTCA 837
 QY 101 AAPCYSPHELEUTHRALLEULEUGLYVLSLEUGILYLSERSPHEANSEVALHISLEU 120
 Db 838 GACTGTTTTCTCACTACTCTGGGTAACCTGAGAAAGAGCTTCAACTCTGTTCACCTG 897
 QY 121 LEUALAGLYGYPHEALAGLUPHESERATCYSPHEPROGLYLEUCYSGILUGLYLSER 140
 Db 898 CTTGACAGTGGGTTTCTGAGTTCTCTCGTTGTTTCCCTGGCTCTGTGAGAGAAATCC 957
 QY 141 THRLEUALPROTHRCYSLISERGINPROCYGLEUPROVALALASNTLEGLYPROTHR 160
 Db 958 ACTCTAGTCCCTACCTGCACTTCTCAGCCTTGCTTACCTTGCCCAACTGGGCCAAC 1017
 QY 161 ARGILELEUPROASNLEUTHYLEUGLYCYSGINARAPVALLEUANLSGLILEU 180
 Db 1018 CGAATTTCTCCCAATCTTATCTTGAGCTCCAGCGAGATGCTCTCAACAAGAGCTGATA 1077
 QY 181 GLINGINASNGLYILEGLYTYRVALLEUANLASERTYRTHRCYSPROLYSPROASPH 200
 Db 1078 CAGCAAGATGGGATTGGTTATGTTAAATGCCAGCTATACCTGTCCAAAGCCTGACTTT 1137
 QY 201 ILEPROGLUSERHISPHLEUARGVALPROVALASAPSERPHECYSGILUGLYLSLEU 220
 Db 1138 ATCCCCGAGTCTCATTTCTGCGTGGCTGTGAAAGACAGCTTTGTGAGAAATTTTG 1197
 QY 221 PROTPLEUASPLYSERVALASPHLEILEGLULYVALALYSALASERASNGLYCYSE 240
 Db 1198 CCGTGTGGACAAATCACTAGTATTTCACTTGAGAAAGCAAAAGCTCCCAATGATGTGT 1257
 QY 241 LEUALHISCYSEUALAGLYILESERATGSEVALATHIRLEALILEALTYRILEW 260
 Db 1258 CTAGTGCATGTTTACCTGGGATCTCCGCTCGGCACCATCGCTATCCCTACATCANG 1317
 QY 261 LYSARGMETASPMESELEUASPGILUALIYTRARGPHEVALYLSGLULYARGPROTHR 280
 Db 1318 AAGAGATGACATGCTTTAGTAGAAGCTTACAGATTGTGAAAAAAGAACCTACT 1377
 QY 281 ILESEPROASNPHESANPHELEUGLYINLEULEUASPTYGLULYSYLSILEYSA 300
 Db 1378 ATRATCCCAAATCTCAATTTCTGGGCCAACTCTGACTATGAGAAAGATTAAAGAC 1437
 QY 301 GINTHRGILYALASERGLYPROLYSERLYLSLEULYLSLEULYLSLEULYSPROAN 320
 Db 1438 CAGACTGGAGCATCAGAGGCCAAAGAACCAACTCAAGCTGTGACTTGAGAAAGCCAAAT 1497
 QY 321 GIUPROVALPROVALVALSERGLUGLYGLINLYSERGLUTHRPROLEUSERPROPO 340
 Db 1498 GAACCTGCTCCCTGCTCTCAGAGGGTGACAGAAAAAGGAGCGCCCTCAGTCCACCC 1557
 QY 341 CYSAIASPSEVALATHRSEGLUALAALAGLYGLINARGPROVALHISPROLASERVAL 360
 Db 1558 TGTGCCGACTCTCTACTCCTCAGAGGACAGAGCAAAAGCCCTGTGATCCCGCAGCG 1617
 QY 361 PROSEVALPROSEVALGINPROSELEULEUGLUAASPERPROLEUVALGINALAEU 380
 Db 1618 CCAGAGGTCCCGAGCTGACAGCGCTGCTTGAAGAGACGCCCTGTGTACAGGCGCTC 1677
 QY 381 SERGLYLEUHHISLEUSERLALAPARGLEUGLUAASPERASULYLSLEULYARGSERPHE 400
 Db 1678 AGTGGGCTCACCTGTGCCAGACAGCTGGAAGACAGCAATAAGCTCAAGCTTCTTC 1737
 QY 401 SERLEUASPIILEYSEVALSERTYRSEVALASERMETALALASERLEUHHISGLYPHE 420
 Db 1738 TCTCTGATATCAAACTGTTTCAATTCAGCCAGATGCGACATCTTACATGCTTC 1797
 QY 421 SERSESEGLUASPALALEUGILYTYRLYSPROSETHRTHIRLEUASPGILYTHRAAN 440
 Db 1798 TCTCATCAGAAAGATGCTTGGAAATCTCAAACTTCCACTGTGATGGAGCAAC 1857

QY 441 LYSLEUCYSGINPHESERPROVALGINGLILEUSERGLUGINTHRPROGLUTHRSEPRO 460
 Db 1858 AAGCTATCCCACTTCTCCCTGTGTCAAGAACTATCGAGAGACTCCGAAACAGTCT 1917
 QY 461 AAPLYSGILUGLUALASERILEPROLYLSLEUGINTHRALARGPROSEASPSERGIN 480
 Db 1918 GATTAAGAGGAAGCCAGCATCCCAAGAAAGCTGACGACCGCAGGCTTTCAGAGCGAG 1977
 QY 481 SERLYSARGLEUHHISSEVALARGTHRSESERSEGLYTHRALAGLARGSERLEU 500
 Db 1978 AGCAAGCATTCATTCGCTGACAAACAGACAGCATGTCACCGCCCAAGGTCCTTTTA 2037
 QY 501 SERPROLEUHHISARGSEGLISERVALGINLUAASPERYRTHIRTHRSEPHLEUPHEGLY 520
 Db 2038 TCTCCATGCAATCGAAGTGGAGCGTGGAGCAATTCACACCAAGCTTCTTTTGGC 2097
 QY 521 LEUSERTHRSEGLINGINHHISLEUTHRILYSESERALAGLYLEUGLYLEULYSGLYTRPHIS 540
 Db 2098 CTTTCCACAGCCACAGACACCTCAAGAGTCTGCTGGCTTGAAGGCTGGCAC 2157
 QY 541 SERASPIILEUALAPROGLINTHRSETHRPROSELEUTHRSESERTRPTYRPHALA 560
 Db 2158 TCGATATCTTGGCCCCCAGACTCTACCCCTTCCAGACAGCAGCTGTGATTTTGC 2217
 QY 561 THRGLUSERSEHISPHETYSERVALASERLALILEYRGLYGLYSERALASERTYRSE 580
 Db 2218 ACAGAGTCTCACATTTCTACTCTCTACGCCATCTACGAGGACATGCTTACTTCT 2277
 QY 581 ALATYRSECYSESEGLINLEUPROTHRCYSGIYASPGINVALYRSEVALARGARG 600
 Db 2278 GCTACAGTGAAGCAGCACTGCGCACTTGCGGAGCAAGTCTATTTGTGGCAGCGG 2337
 QY 601 GLINLYSPROSEARAPARALASPSERARGARGSETRPHIAGLUGLUSERPROHEGLU 620
 Db 2338 CAGAAAGCCAGAGAGAGCTGACCTGCGGAGCACTGGCATGAAGAGAGCCCTTTGAA 2397
 QY 621 LYSGLINPHELYSARGSERCYSGINMETGLUPHEGLYGLUSERILEMETSEGLUASN 640
 Db 2398 AAGCATTTAAACGAGAAAGCTGCCAAATGGAATTTGAGAGACATCATGACAGAAC 2457
 QY 641 ARGSEARGJUGLULEUGLYLYSVALGLYSEGINSESPHESEGLYSEMETG 660
 Db 2458 AGGTACCGGAAAGACTGGGGAAGTGGGAGTCAGTCTAGCTTTTCCGGCAGCATGAA 2517
 QY 661 ILEILEGLUVALSER 665
 Db 2518 ATCATTTGAGTCTCC 2532
 RESULT 3
 ID ADA53105 standard cdNA, 2102 BP.
 AC ADA53105;
 DT 20-NOV-2003 (first entry)
 DE Human coding sequence, SEQ ID 673.
 OS Homo sapiens.
 PN EP1293569-A2.
 PD 19-MAR-2003.
 PF 21-MAR-2002; 2002EP-00006586.
 PR 14-SEP-2001; 2001JP-00328381.
 PR 24-JAN-2002; 2002US-0350435P.

XX (HELI-) HELIX RES INST.
 PA (REAS-) RES ASSOC BIOTECHNOLOGY.
 XX
 PI Isogai T, Sugiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S,
 PI Yamamoto J, Isono Y, Hio Y, Otsuka K, Nagai K, Irie R, Tamechika I,
 PI Seki N, Yoshikawa T, Otsuka M, Nagahara K, Masuno Y;
 XX
 DR WPI; 2003-395539/38.
 DR P-PSDB; ADA54744.
 XX
 PT New polynucleotides encoding full-length polypeptides, e.g. secretory
 PT and/or membrane proteins, useful for developing medicines for diseases in
 PT which the gene is involved, or as target molecules for gene therapy.
 XX
 PS Claim 1; SEQ ID NO 673; 205bp; English.
 XX
 CC The present invention relates to novel human secretory or membrane
 CC proteins (ADA54072-ADA55710) and their coding sequences (ADA52433-
 CC ADA54071). The coding sequences are useful in the gene therapy of
 CC diseases caused by abnormalities of the proteins, e.g. cancer,
 CC inflammatory diseases, osteoporosis or neurological disease.
 CC
 XX
 SQ Sequence 2102 BP; 542 A; 546 C; 510 G; 504 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0	Length:	2102
Score:	472.00	Matches:	662
Percent Similarity:	99.10%	Conservative:	0
Best Local Similarity:	99.10%	Mismatches:	3
Query Match:	70.98%	Indels:	6
DB:	7	Gaps:	0

US-10-029-345A-109 (1-665) x ADA53105 (1-2102)

QY 1 MetAlaHisGluMetIleGlyThrGlnIleValThrGluArgLeuValAlaLeuLeuGlu 20
 DB 56 ATGGCCCATGAGTATGATGAACTCAATTGTTACTGAGAGTGGTGGCTGCTGGAA 115
 QY 21 SerGlyThrGluValLeuLeuIleAspSerArgProPheValGlnTyrAsnThrSer 40
 DB 116 AGTGAACCGAAGAAAGTGTGCTAATTGATAGCCGCGCATTTGTGMAATACATCATCC 175
 QY 41 HisIleLeuGluAlaIleAsnIleAsnGlySerIleLeuMetIleArgAlaGlnGln 60
 DB 176 CACATTTTGGAAACCATTAATATCACTGCTCAAGCTTAAGAAGCAAGTTGCAACAG 235
 QY 61 AspLeuValLeuIleThrGluLeuIleGlnHisSerAlaIleHisIleValAspIleAsp 80
 DB 236 GACAAAGTGTTAATTACAGAGCTCATCCAGCATTCAGCGAAACATTAAGTTGACATTGAT 295
 QY 81 CysSerGlnIleValValValTyrAspGlnSerSerGlnAspValAlaSerLeuSerSer 100
 DB 296 TGCAGTCAGAAAGTTGATGATTACATCAAGTCCCAAGATGTTGCTCTCTCTTCA 355
 QY 101 AspCysPheLeuThrValLeuLeuGlyValLeuGluIleSerPheAsnSerValHisLeu 120
 DB 356 GACTGTTTTCTCACTGACTTCTGGGTAACTGAGAGAGACTTCAACTCTGTTCACTG 415
 QY 121 LeuAlaGlyValPheAlaGluPheSerArgCysPheProGlyLeuGlyGluIleValSer 140
 DB 416 CTTCGAGGTGGGTTTGTGAGTTCTCTCGTTGTTTCCCTGAGCTCTGTGAAGAAATCC 475
 QY 141 ThrLeuValProThrCysIleSerGlnProCysLeuProValAlaAsnIleGlyProThr 160
 DB 476 ACTCTAGTCCCTACCTGATCTTCTCAGCCTTGCTTACCTGTTCCAACTGGGCCAAC 535
 QY 161 Arg-IleLeuProAsnLeuTyrIleGlyCysGlnArgAspValLeuAsnIleGluLeu 180
 DB 536 CT-AATTCCTCCAACTTATCTTGTGCTGCGAGCGAGATGCTCAACAAGAGCTGAT 594
 QY 180 e-GlnGlnAsnGlyIleGlyTyrValLeuAsnAlaSerTyr-ThrCysProLysProAsp 199
 |||||||

DB 595 -CGACGAAATGGAGTTGGTTATGTATTATGCCAGCAA-TACCTGTCCAAAGCTGAC 652
 QY 200 PheIleProGluSerHisPheLeuArgValProValAsnAspSerPheCysGluValIle 219
 DB 653 TTTATCCCGAATCTCATTTCTCGGTGTCCTGGTGAAGACAGCTTTGTGGAATAAT 712
 QY 220 LeuProThrLeuAspIleSerValAspPheIleGluValAlaValAsnAsnGlyCys 239
 DB 713 TTGCGGTGTTGGCAAAATCAGTAGATTTCATTGAGAAAGCAAAAGCTTCCATGATGT 772
 QY 240 ValLeuValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIle 259
 DB 773 GTTCTAGTCACTGTTTACGTGGATCTCCCGCTCGGACCATTCGCTATCGCTACATC 832
 QY 260 MetIleValMetAspMetSerLeuAspGluAlaIleTyrArgPheValIleGluValAspPro 279
 DB 833 ATGAAGAGATGACATGTCTTAGTAGAAGTTGACAGTTGTGMAAAMAAACCACT 892
 QY 280 ThrIleSerProAsnPheAsnPheLeuGlyGlnLeuLeuAspTyrGluValValIleVal 299
 DB 893 ACTATATCTCCAACTTCAATTTTCTGGGCCAACTCTTGACTATGAAAGAAATTAAG 952
 QY 300 AsnGlnThrGlyAlaSerGlyProLysSerIleValValLeuLeuHisIleGluIleValPro 319
 DB 953 AACCAAGACTGAGCATACAGGCGCAAGCAAACTCAACTGCTGACCTGGAGAGCCCA 1012
 QY 320 AsnGluProValProAlaValSerGluGlyGlnIleValSerGluThrProLeuSerPro 339
 DB 1013 AATGAACCTGTCTCTGCTCTCTCAGAGGTGACAGAAAGAGACGCGCCCTCACTCCA 1072
 QY 340 ProCysAlaAspSerAlaThrSerGluAlaIleGlyGlnArgProValHisProAlaSer 359
 DB 1073 CCTGTGCGACTGTGTAAGTCTCAGAGCGAGCAAGCAAAAGCCGTGATCCCGCAGC 1132
 QY 360 ValProSerValProSerValGlnProSerLeuLeuGluAspSerProLeuValGlnAla 379
 DB 1133 GTGCCAGGCTGCCAGCGCTGACGCGTGTGTTAGAGACAGCCGCTGGATACAGGGG 1192
 QY 380 LeuSerGlyLeuHisIleSerAlaAspArgLeuGluAspSerAsnValLeuValAspSer 399
 DB 1193 CTCAGTGGGCTCACCCTGTCCGACAGCGCTGCAAGACAGCAATTAAGCTCAAGCCTTCC 1252
 QY 400 PheSerLeuAspIleLeuSerValSerTyrSerAlaSerMetAlaIleSerLeuHisGly 419
 DB 1253 TTCTCTGATATCAAAATCAGTTTCAITTAAGCCAGATGACAGCATCTTACATGGC 1312
 QY 420 PheSerSerSerGluAspAlaLeuGluTyrTyrIleProSerThrThrLeuAspGlyThr 439
 DB 1313 TTCTCTCATCAGAAAGATGCTTTGGAAATCTCAAAACCTTCCACTACTGTGATGGAGC 1372
 QY 440 AsnValLeuCysGlnPheSerProValGlnIleLeuSerGluGlnThrProGluThrSer 459
 DB 1373 AACCAAGCTATGCAAGTTCTCCCTGTTCAAGAACTTCGAGAGACAGCTCCGCAACCAAGT 1432
 QY 460 ProAspValGluGluAlaSerIleProValValLeuGlnThrAlaArgProSerAspSer 479
 DB 1433 CCTGATTAAGAGAAAGCCAGCATCCCAAGAAAGCTGCAAGCTCCAGGCTTCAAGCAC 1492
 QY 480 GlnSerValArgLeuHisSerValArgThrSerSerSerGlyThrAlaGlnArgSerLeu 499
 DB 1493 CAGAGCAACGATTTGATTCGATTCGATCAGAACACAGACAGAGTGCACCGCCAGAGTCCCTT 1552
 QY 500 LeuSerProLeuHisArgSerGlySerValGluAspAsnTyrHisThrSerPheLeuPhe 519
 DB 1553 TTATCTCCACTCATCAAGTGGAGCGTGGAGGCAATTAACACACAGCTTCTCTTTC 1612
 QY 520 GlyLeuSerThrSerGlnIleHisLeuThrIleSerAlaGlyLeuGlyIleValSerGlyTyrP 539
 DB 1613 GGCCTTTCCACAGCAGAGCACTTCAGAAAGTGTGCTGGGCTTGAAGGGCTGG 1672
 QY 540 HisSerAspIleLeuAlaProGlnThrSerThrProSerLeuThrSerSerTyrTyrPhe 559
 DB 1673 CACTGGATATCTTGGCCCCCAGACCTTACCCCTTCCCTACACAGACGCTGATTTT 1732

QY 560 AlathrgIuSerSerHisphenTySerAlaSerAlaIleTyrgIyglYserAlaSerTyr 579
 DB 1733 GCCACAGAGTCTTACACTTCTACTCTGCTTCAGCCATCTACGGAGGCACTGGCAGTTAC 1792
 QY 580 SerAlaTySerCySerSerGlnLeuProThrCysGlyAspGlnValTySerValArgArg 599
 DB 1793 TCTGCTTACAGCTCCAGCCAGCTGCCCATTTGGCGAACAACAAGTCTATTCTTGCGCCAG 1852
 QY 600 ArgGlnTyProSerArpaAlaAspSerArgArgSerTphIsgIuSerProPhe 619
 DB 1853 CGGAGAGAGCCAAATGACAGAGCTGACTCCGCGGAGCTGGCATGAGAGAGACCCCTTT 1912
 QY 620 GluTySerGlnPheTySArGArSerCySGlnMetGlnPheGlyglYserIleMetSerGln 639
 DB 1913 GAAAGAGAGTTTAAACCAAGAGCTGCCAAATGAAATTGAGAGAGCATCATGTCAAG 1972
 QY 640 AsnArgSerArgGlnIuSerGlyTyValGlySerGlnSerSerPheSerGlySerMet 659
 DB 1973 AACAGGTCACGGAGAGAGCTGGGGAAAGTGGGCAAGTCACTTTCGGGCAAGCATG 2032
 QY 660 GluIleIleGluValSer 665
 DB 2033 GAAATCATTTAGGCTCTCC 2050

RESULT 4
 AAD09492
 ID AAD09492 standard; DNA; 2732 BP.

AC AAD09492;
 DT 10-SEP-2001 (first entry)
 DE Human SGP002 phosphatase polypeptide encoding DNA.

XX Human; SGP002 phosphatase polypeptide; phosphatase-related disease;
 KW immune-related disorder; ocular disease; organ transplant rejection;
 KW infection; diabetes; pain; sexual dysfunction; Alzheimer's disease;
 KW metabolic disorder; haematopoietic cancer; mood disorder; cardiac;
 KW Parkinson's disease; multiple sclerosis; amyotrophic lateral sclerosis;
 KW cardiovascular disease; brain; neuronal-associated disease; dyskinesia;
 KW attention disorder; cognition disorder; psychotic disorder; cystostatic;
 KW neurological disorder; vitruide; noctropic; cerebroprotective; therapy;
 KW neuroprotective; antibacterial; vulnery; cranialliser; antiaesthetic;
 KW hypotensive; immunosuppressive; antiporiatic; analgesic; hyperensive;
 KW antifungal; dual specificity phosphatase; DSP; MAP kinase phosphatase;
 KW MKP; migraine; chromosome 12p11.1-p12.1; ds.

XX Homo sapiens.

OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 538..2535
 FT /tag= a
 FT /product= "Human SGP002 phosphatase polypeptide"

PN MO200146394-A2.

XX 28-JUN-2001.

PD 21-DEC-2000; 2000MO-US034736.

PF 21-DEC-1999; 99US-0172255P.

PR 28-DEC-1999; 99US-0175766P.

PR 25-JAN-2000; 2000US-0178078P.

PR 31-JAN-2000; 2000US-0179301P.

XX (SUGB-) SUGEN INC.

XX Plowman GD, Martinez R, Whyte D, Manning G, Sudareanam S;
 PI Hill RJ, Flanagan P;
 XX WPI: 2001-418058/44.
 DR P-PSDB; AAE04834.

XX Novel phosphatase polypeptide useful for treating cancers, immune-related
 PT diseases and disorders, cardiovascular disease, brain or neuronal-
 PT associated diseases and metabolic disorders.
 XX
 PS Claim 29; Fig 1; 186pp; English.

CC The present invention relates to phosphatase polypeptides, nucleotide
 CC sequences encoding them, as well as various products and methods useful
 CC for the diagnosis and treatment of various phosphatase-related diseases
 CC and conditions. Substance that modulates the activity of phosphatase
 CC polypeptide is used to treat immune-related diseases and disorders,
 CC cardiovascular disease, brain or neuronal-associated diseases and
 CC metabolic disorders, including cancers of tissues, cancers of
 CC haematopoietic origin, diseases of central and peripheral nervous system,
 CC Alzheimer's disease, Parkinson's disease, multiple sclerosis, amyotrophic
 CC lateral sclerosis, viral infections, infections caused by prions,
 CC bacteria and fungi, ocular diseases, diabetes, migraines, pain, sexual
 CC dysfunction, mood disorders, attention disorders, cognition disorders,
 CC hypertension, psychotic disorders, neurological disorders,
 CC dyskinesias and organ transplant rejection. The present sequence is a DNA
 CC encoding human SGP002 phosphatase polypeptide. This sequence is
 CC classified as dual specificity phosphatase (DSP) and MAP kinase
 CC phosphatase (MKP). SGP002 gene maps to chromosomal position 12p11.1-p12.1

XX Sequence 2732 BP; 710 A; 684 C; 686 G; 652 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0 Length: 2732
 Score: 472.00 Matches: 663
 Percent Similarity: 99.40% Conservative: 0
 Best Local Similarity: 99.40% Mismatches: 2
 Query Match: 70.98% Indels: 4
 DB: 4 Gaps: 0

US-10-029-345A-109 (1-665) x AAD09492 (1-2732)

QY 1 MetAlaHisGluMetIleGlyThrGlnIleValThrGluArgLeuValAlaLeuLeuGlu 20
 DB 538 ATGGCCCAAGTGAAGATTTGAAGCACTCAAAATGTTACTAGAGAGTTGGCTGTGCGAA 597
 QY 21 SerGlyThrGluValLeuLeuIleAspSerArgProPheValGluTyAsnThrSer 40
 DB 598 AGTGAAGCGAAAGAGCTGCTTAATGATAGCCGCCATTGGGAATCAATACATCC 657
 QY 41 HisIleLeuGluAlaIleAsnIleAsnCySerIleuMetIleuTyArgArgLeuGln 60
 DB 658 CACATTTGGAAGCCATTAAATCACTGCTCCAGCTTATGAAGCGAAGTTGCAACAG 717
 QY 61 AspIleValLeuIleThrGluLeuIleGlnHisSerIleValHisIleValAsp 80
 DB 718 GACAAAGTGTAAATTTACAGAGCTCATCCAGCATTCAGCGAAACATTAAGTTGACATTGAT 777
 QY 81 CysSerGlnIleValValIleTyAspGlnSerSerGlnAspValAlaSerLeuSerSer 100
 DB 778 TGCAGTCAGAAAGTGTAGTTTACGATCAAAAGCTCCCAAGATGTGCTCTCTCTCA 837
 QY 101 AspCySPheIleuThrValIleLeuLeuGlyValSerPheAsnSerValHisLeu 120
 DB 838 GACTGTTTCTCACTGATCTTGGGTAACTGGAGAAAGACTTCAACTCTGTTCACTG 897
 QY 121 LeuAlaGlyIlePheAlaGluPheSerArgCysPheProGlyLeuCySGluIleValSer 140
 DB 898 CTTCGAGGTGGTGTGCTGAGTTCTCTCGTTGTTTCCCTGCGCTCTGTGAAGAAATCC 957
 QY 141 ThrLeuValProThrCysIleSerGlnProCysLeuProValAlaAsnIleGlyProThr 160
 DB 958 ACTCTAGTCCCTACCTGACATTTCTCAGCCTTGCTTACTGTTCCAAACATGGGCCAAC 1017
 QY 161 ArgIleLeuProAsnIleuTyLeuGlyCySGlnArgAspValLeuAsnIleGluLeu 180
 DB 1018 GGAATCTTCCCAATCTTATCTTGGCTGCCAGAGATGTCTCTCAACAGAGCTGAT- 1076

QY	181	GLNGLNANGNGLYEGLYTTRVVALLEUNASNAASERLYR-ThrCybProLysPProAaP	200
Db	1077	GCAGCAAGATGGGATGGTTATGTGTAAATCCAGCAA-TACCTGTCCAAAGCCTGACT	1139
QY	200	heLEProGLuSerHisPheLeuAArgValProValASnaAPSerPheCySGluYSIleL	220
Db	1136	TTATCCCGGAGTCTATTTCCTGGCTGTGCCGTGTGAATGACAGCTTTGTGAGAAATTT	1195
QY	220	eUProCTrPLeuAspLysSerValAaPPhelLEGluValAlaYSaAsSerANGLYCybV	240
Db	1196	TGCCCTGGTTGGACAAATCAGTAGATTTTCATTGAGAAAGCAAAAGCTCCATGTGATGTG	1255
QY	240	alleuValHisGlyLeuAlaGlyLIIeserAArgSerHATHrIleAlaIleAlaTYrIleM	260
Db	1256	TTCTATGTCCACTGTTTACCTGGGATCTCCCGCTCCGCAACATCGCTATCGCTACATCA	1315
QY	260	eLYAArgMeCAsPMeSerLieuAspGluAlaTYArgPheValLYSGluYSaArgProT	280
Db	1316	TGAGAGGATGGACATGTCTTTAGTGAAGCTTACAGATTTGTGAAAGAAAAGAAAGCTTA	1375
QY	280	hriLEserProAaNPheAaNPheLeuGlyGluLeuLeuAspTYrGluYSaYSIleLYa	300
Db	1376	CTATATCTCCAAACTTCATTATTTCTGGGCAACTCTGGACATAGAGAAGAAATTAGA	1435
QY	300	enGlnThrGLYAlaSerGlyProLYSerLYSLeuLYSLeuLHISLeuLYSProA	320
Db	1436	ACCAACACTGGACATCAGGCGCCAAAGAGCAAACTAGCTGTGCACCTGGAGAGCCAA	1495
QY	320	enGluProValProAlaValserGluGlyGluLYSerGluThrProLeuSerProP	340
Db	1496	ATGAACCTGTCCCTCTGTCTCAGAGGTTGACAAAGAAAGCAGCGCCCTCAGTCCAC	1555
QY	340	roCYaLaAspSerAlaThrSerGluAlaAlaGlyGlnArgProValHisProAlaSerV	360
Db	1556	CCTGTGCGGACCTCTGACTACCTCAGAGGACAGAGCAAAAGCCGTGATCCCCGACAG	1615
QY	360	aLProSerValProSerValGlnProSerLeuLeuGluAspSerProLeuValGlnAlaL	380
Db	1616	TGCCCAGGCTGCCCGCGTGCAGCCGTCTGTGTAGAGACAGCCCGCTGTACAGCGGC	1675
QY	380	eUSerGlyLeuLHISLeuSerAlaAaPAArgLeuGluAspSerAsnLYSLeuYSaArgSerP	400
Db	1676	TCAGTGGCTGCACCTGTCCGACAGAGGCTGGAAAGACAGCAATAGCTCAAGCCTTCT	1735
QY	400	heSerLeuAspLIElySerSerValSerTYSerAlaSerMechAlaAsSerLeuHISGlyP	420
Db	1736	TCTCTCTGGATATCAATCAATCAAGTTTCAATTCAGCCAGCATGGCAGCATCTTACATGCT	1795
QY	420	heSerSerSerGluAspAlaLeuGluTYrTYrLYSProSerThrThrLeuAspGlyThra	440
Db	1796	TCTCTCTATCAGAAATGATCTTTGGAAATCTACAAACCTTCCACTCTGATGGGACCA	1855
QY	440	enLYSLeuCYGlnInheSerProValGlnGluLeuSerGluGlnInhrProGluThrSerP	460
Db	1856	ACAAAGCTATGCCAGTCTTCCCTGTTCAGGAACTATGGAGAGAGCTCCGAAACAGCTC	1915
QY	460	roAspLYSGluGluAlaSerLIEProLYSLeuGlnThrAlaArgProSerAspSerG	480
Db	1916	CTGAATAGAGAGAACCCAGCATCTCCCAAGAGCTCAGACTCCAGGCTTTCAGACAGCC	1975
QY	480	InSerLYaArgLeuHISSerValArgThrSerSerSerGlyThraAlaGlnArgSerLeuL	500
Db	1976	AGAGCAACGATTCGATTCGTCAGAAACAGCAGCAGTGGCACCCCCAGAGGTCCCTTT	2035
QY	500	eUSerProLeuHISArgSerGlySerValGluAspAsnTYrHisPheLeuPheG	520
Db	2036	TATTCCTCACTGATGAAGTGGAGACCGTGGAGAGAAATTTACACACAGCTTCCTTTTCG	2095
QY	520	LYLeuSerThrSerGlnGlnHISLeuThrLYSLeuAlaGlyLeuGlyLYSLeuLYSArgP	540
Db	2096	GCTTTTCCACCGCAAGCAGCACTCAGCAAGTCTGTGGCTGGGCTTAAAGGCTGGC	2155
QY	540	ISerAspLIEleuAlaProGlnInhrSerThrProSerLeuThrSerSerTrpTYrPheA	560

Dd		2156	ACTGGATATCTTGTGACCCCAAGACCTCTACCCCTTCCTGCACGACAGCTGGATTTTTG	2215
Oy		560	IATHRGUSESESRHAPHETYISERXALASERXALLEYRGJYGJYSERXASERTYRS	580
Dd		2216	CCACAGAGCTCCTCACATTTCTACTCTGCGTAGCATCTACCGAGGCGAGGCCAGTTACT	2275
Oy		580	ERLAIATYRSECYSESERGILNEUPROTHRCYSGEIJYABGINVALTYRSEVALARGA	600
Dd		2276	CTGGCTTACAGCTGACGACGAGCTGGCCACTTGGCGGAGCCAAAGCTATTTCTGTGCGCAGGC	2335
Oy		600	RGLINLVPRPSERARSPARIALASPSEXRARGATRGSETRPHJSGIUGIUSERPROPHG	620
Dd		2336	GCGAGAACCCAGAGACAGAGCTGACTCTGGCGGAGACTGCAATGAAGAAGGCCCTTTG	2395
Oy		620	IULVSGINPHELRYARGARSESCYSGSINMETGIUHPNEGJIYGIUSERILEMESERGUNA	640
Dd		2396	AAAAGCATTTTAACGACGAAAGCTGCCCAATTGAAATTTGAGAGAGCATCATGTACAGAGA	2455
Oy		640	SNAVGSEARVGLUGLIULENGLYLYVAIGLYSERGINSESRPHESERGJYSERWEG	660
Dd		2456	ACAGGTGACCGGGAAGACTGGGGGAATACTGGCGAGTCAGTCTACCTTTTCGGGACGATCG	2515
Oy		660	IUIELLEGLUVASER	665
Dd		2516	AAATCATTTGAGTCTCC	2532
DE	RESULT 5			
XX	AAH99685		AAH99685 standard; cDNA; 2966 BP.	
XX	AC	AAH99685;		
XX	DT	16-OCT-2001	(first entry)	
DE	Human protein encoding cDNA sequence SEQ ID NO:520.			
KM	Human; cancer; ulcer; HIV infection; human immunodeficiency virus;			
KM	antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;			
KM	antibacterial; endocrine; cardiac; central nervous system; vitruide;			
KM	anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;			
KM	antiggregant; haemostatic; vulnery; antitumor; osteopathic; eczema;			
KM	dermatological; antiallergic; antiaesthetic; antididiabetic; cycostatic;			
KM	neuroprotective; antidepressant; nootropic; antiParkinsonian; infection;			
KM	immunostimulant; gene therapy; antilease therapy; vaccine; inflammation;			
KM	antianaphylactic; rheumatoid arthritis; septic shock; pancreaticitis;			
KM	cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;			
KM	genetic dyspeasia; haematopoietic disorder; platelet disorder; actinia;			
KM	thrombocytopoenia; osteoporosis; severe combined immunodeficiency;			
KM	allergic rhinitis; diabetes; multiple sclerosis; depression;			
KM	Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;			
KM	neurological disorder; ss.			
OS	Homo sapiens.			
PN	WO200153455-A2.			
PD	26-JUL-2001.			
PF	22-DEC-2000; 2000WO-US035017.			
PR	23-DEC-1999; 99US-00471275.			
PR	21-JAN-2000; 2000US-0048725.			
PR	25-APR-2000; 2000US-00552317.			
PA	(HYSE-) HYSEQ INC.			
PI	Tang YT, Liu C, Drmanac RT;			
DR	WPL: 2001-457603/49.			
PT	P-PBDB; AAM25744.			
PT	Isolated human polynucleotides encoding polypeptides, useful for the			

PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection.

XX Claim 1, Page 578, 1217pp, English.

XX
CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to
CC AAM25963. The proteins can have activities based on the tissues and cells
CC they are expressed in, such as: antiinflammatory; antineumatic;
CC antiallergic; immunosuppressive; antibacterial; endocrine; cardiac;
CC central nervous system; virocid; anti-HIV; fungicide; antimutagen;
CC cardiovascular; antianaemic; antiagregant; haemostatic; vulnerray;
CC anticancer; osteopathic; dermatological; antiallergic; antistatic;
CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
CC encoding them can be used in gene therapy, antisense therapy and vaccine
CC production. The proteins and polynucleotides are useful for screening for
CC of disorders associated with the activity of a protein and for the treatment and diagnosis
CC of disorders associated with the activity of a protein e.g. inflammation,
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
CC anaemia, platelet disorders, thrombocytopenia, wounds, burns, ulcers,
CC osteoporosis, severe combined immunodeficiency, eczema, allergic
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
CC neurological disorders

XX Sequence 2966 BP; 809 A; 735 C; 688 G; 733 T; 0 U; 1 Other;

Alignment Scores:

Pred. No.: 0
Score: 472.00
Percent Similarity: 99.39%
Best Local Similarity: 99.39%
Query Match: 70.98%

Length: 2966
Matches: 648
Conservative: 0
Mismatch: 2
Indels: 4
Gaps: 0

US-10-029-345A-109 (1-665) x AAH99685 (1-2966)

QY 16 ValAlaLeuLeuGluSerGlyThrGluValLeuLeuLeuAspSerArgProPheVal 35
DB 68 GGGGCTCTCTGGAAAGTGGAAAGGAAAGTCTGTAATGATGATGCGGCAATTTGTG 127
QY 36 GluTYrAsnThrSerHisLeuGluAlaIleAsnIleAsnGlySerLysLeuMetLys 55
DB 128 GAATCAATATCATCCACATTTTGGAAAGCATTAAATATCACTGCTCCAAAGTTATGAAG 187
QY 56 ArgArgLeuGlnGlnAspLysValLeuLeuThrGluLeuLeuIleGlnHisSerAlaLysHis 75
DB 188 CGAAGGTTCACACAGACAAAGGTGTTAATTACAGAGCTCATCCACATTCACGCAAAACAT 247
QY 76 LysValAspIleAspCysSerGlnLysValValValTYrAspGlnSerSerGlnAspVal 95
DB 248 AAGGTTGAATGATTGATTCAGTCAAGAAAGTTGATTTACGATCAAAAGCTCCCAAGATGT 307
QY 96 AlaSerLeuSerSerAspCysPheLeuThrValLeuLeuGlyLysLeuGlnLysSerPhe 115
DB 308 GCTCTCTCTCTTCAGACTGTTTCTCACTGACTTCTGGGTAACCTGGAAGAGCTTC 367
QY 116 AsnSerValHisLeuLeuValGlyGlyPheAlaGlnPheSerArgCysPheProGlyLeu 135
DB 368 AACTCTGTTACCTGCTGGAGGTGGGTGCTGAGGTCTCTGTTGTTTCCCTGGCTTC 427
QY 136 CysGlnGlyLysSerThrLeuValProThrCysIleSerGlnProCysLeuProValAla 155
DB 428 TGTGAAGAAATTCACACTCTGTCCTTACCTGCAATTTCTCAAGCTTCTTACCTGTTCC 487
QY 156 AsnIleGlyProThrArgIleLeuProAsnLeuTYrLeuGlyCysGlnArgAspValLeu 175
DB 488 AACATTGGGCCCAACCGCAATCTTCCCAATCTTTATCTTGCTGCACAGAGATGCTCTC 547
QY 176 AsnLysGluLeuIle-GlnGlnAsnGlyIleGlyTYrValLeuAsnAlaSerTYr-ThrC 195
DB 548 AACAGAGAGCTGAT-GCAGCAGAAATGGGATTGTATGTATGTTAATGCCAGCAA-TACCT 605

QY 195 YspProLysProAspPheIleProGluSerHisPheLeuArgValProValAsnAspSerP 215
DB 606 GTCCAAAGCCCTGACCTTATCCCGAGTCTCAATTCCTCGTGTGCTGTGAATGACAGCT 665
QY 215 heCysGlyLysIleLeuProThrLeuAspLysSerValAspPheIleGlnLysValAlaLysA 235
DB 666 TTTTGGAGAAATTTTGGCCGTGGTGGACAAATCAGTGAATTTCTTATGAAAGCAAAAG 725
QY 235 LaseAsnGlyCysValLeuValHisCysLeuAlaGlyIleSerArgSerAlaThrIleA 255
DB 726 CCTCAAGATGATGTTCTAGACACTGTTAGCTGGGATCTCCCGCTCCGCAACATTCG 785
QY 255 LaIleAlaTYrIleMetLysArgMetLaspMetSerLeuAspGluAlaTYrArgPheVal 275
DB 786 CTATGCTTACATCATGAAAGAGATGACATGCTTTAGATGAAGCTTACAGATTTTGA 845
QY 275 YsgLulYArgProThrIleSerProAsnPheAsnPheLeuGlyGlnLeuAspTYrG 295
DB 846 AAGAAAAAGACCTATATCTCCAAACTTCAATTTTCTGGCCAACTCTGGAATATG 905
QY 295 LulYsLysIleLysAsnGlnThrGlyLaseGlyProLysSerLysLeuLysLeuL 315
DB 906 AAGAGAAATTAAGAACACAGACTGAGACATGAGGCAAGAGCAAACTCAAGCTGCTGC 965
QY 315 LsLeuGluLysProAsnGluProValProValAlaLaseGluGlyGlnLysSerGluT 335
DB 966 AACTGAGAAAGCCAAATGAACCTTCTCTGCTGCTCAAGAGGTGACAGAAAGCCAGCA 1025
QY 335 hrProLeuSerProProCysAlaAspSerAlaThrSerGluAlaAlaGlyGlnArgProV 355
DB 1026 CGCCCTTCAAGTCCACCTGTGCGCACTGTGTACTCAGAGGACAGACAAAGGCCCG 1085
QY 355 AlHisProAlaSerValProSerValProSerValGlnProSerLeuLeuLysSerP 375
DB 1086 TGCATCCCGCCAGGCTGCCAGCTGCGCCAGCGTGCAGCCGTGCTTGAAGACAGACC 1145
QY 375 LoleuValGlnAlaLeuSerGlyLeuHisLeuSerAlaAspArgLeuGlnAspSerL 395
DB 1146 CGCTGATACAGGCCCTCACTGAGGCTGCACTGTCCGACAGAGGCTGGAAGACGCAATA 1205
QY 395 YsLeuLysArgSerPheSerLeuAspIleLysSerValSerTYrSerAlaSerMetAla 415
DB 1206 AGCTCAAGCGTCTCTCTCTGGAATATCAATAGTTTATATTCAGCCAGATGGCAG 1265
QY 415 LaseLeuHisGlyPheSerSerSerGlnAspAlaLeuGluTYrTYrLysProSerThr 435
DB 1266 CATCTTACATGGCTCTCTCTCATCAGAAAGATGCTTGGAAATCTACAAACCTTCACTA 1325
QY 435 hrLeuAspGlyThrAsnLysLeuCysGlnPheSerProValGlnGluLeuSerGluGlnT 455
DB 1326 CTCTGAGTGGACCAACAGCTATGCGAGTCTCCCTGTTCAGAACTATCGAGAGCA 1385
QY 455 hrProGluThrSerProAspLysGlnGluLaseSerIleProLysLysLeuGlnThrAla 475
DB 1386 CTCCGAAACACAGTCCGATTAAGAGSAAAGCCAGCANTCCCAAGAGCTCACAACGCCA 1445
QY 475 rgrProSerAspSerGlnSerLysArgLeuHisSerValArgThrSerSerSerGlyThra 495
DB 1446 GGCCTTCAGACACACAGACAGACGATGTCATTCGGTCAGAACACACAGACAGTGCACCG 1505
QY 495 LagaLysSerLeuLeuSerProLeuHisArgSerGlySerValGlnAspAsnTYrHisT 515
DB 1506 CCAAGAGTCCCTTTTATCTTCACTGCAATGAGTGGAGCGGAGAGAAATTCACACA 1565
QY 515 hrSerPheLeuPheGlyLeuSerThrSerGlnGlnHisLeuThrLysSerAlaGlyLeuG 535
DB 1566 CCAAGCTTCTTTTCGGCTTTCACACAGCAGACAGACCTCAAGAGTGTGCTGGCTGG 1625
QY 535 LysLeuLysGlyTYrPheLaseAspIleLeuAlaProGlnThrSerThrProSerLeuThs 555
DB 1626 GCCTTAAGGAGCTGCACTGGATATCTTGGCCCCCAGACCTTACCTTCCCTGACACA 1685

QY 555 exSerTrpIyrPhealathrGluSerSerHisPheTrYSerAlaSerAlaIleTyrglyG 575
 Db 1686 GCAAGCTGATTTTGGCAGAGAGTCTCAACACTTACTGTGCTCAGCCATCTAGGAG 1745
 QY 575 lYSerAlaSerTrYSerAlaTySerCySserGlnLeuProThrCySglYAspGlnValr 595
 Db 1746 GCAGTCCAGTTACTGTGCTTACAGTGCAGCCAGCTGCCCACTTGCGAGACCAAGTCT 1805
 QY 595 yfSerValArgaArgGlnIlyeProSerSerPargAlaAspSerArgaArgSerTrpHisG 615
 Db 1806 ATTCTGTCGACAGCGCAGAACCCAGAGTACAGAGTCTGCGCGAGCTGCGCATG 1865
 QY 615 lUGluSerProPheGlnIlyeGlnPheIlyeArgaArgSerCySglMetGlnPheGlyGln 635
 Db 1866 AAGAGAGCCCTTTGAAAGACATTTAAAGCAGAACCTGCCAAATGGAATTTGGAGAGA 1925
 QY 635 erlMetSerGluAmaArgSerArgGluGlnIlyeValIglYSerGlnSerSerP 655
 Db 1926 GCATCATGTCAAGAACAGGTCAAGGAGAGCTGGGGAAGTGGCAGCATCTAGCT 1985
 QY 655 heSerGlySerMetGluIleIleGluValSer 665
 Db 1986 TTTCCGGCAGCATGAAATCATTAGGCTCTCC 2017

RESULT 6
 AAS15768
 ID AAS15768 standard; cDNA; 3059 BP.
 AC AAS15768;
 AC AAS15768;
 DT 18-JUN-2002 (first entry)
 XX
 DE cDNA sequence of human (dual specificity phosphatase) DUSP-10.
 KW Human; dual specificity phosphatase; DUSP-10; cancer; epilepsy; stroke;
 KW neuronal degeneration syndrome; Alzheimer's disease; depression;
 KW schizophrenia; asthma; immune disorder; inflammatory process; arthritis;
 KW osteoporosis; diabetes; gene therapy; chromosome 12; 86.
 XX
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 127..2124
 FT /*tag a
 FT /product= "DUSP-10 protein"
 XX
 PN WO200177340-A1.
 XX
 PD 18-OCT-2001.
 XX
 PF 06-APR-2001, 2001MO-EP003966.
 XX
 PR 10-APR-2000; 2000EP-00107143.
 XX
 PA (MERE) MERCK PATENT GMBH.
 XX
 PI Duecker K;
 XX
 DR WPI; 2002-010917/01.
 DR P-PSDB; AAU09946.
 XX
 PT Novel dual specificity phosphatase polypeptides useful for treating
 PT cancer, epilepsy, stroke, neuronal degeneration syndromes, Alzheimer's
 PT disease, depression, schizophrenia, asthma and immune disorders.
 XX
 PS Claim 5; Page 34-37; 43pp; English.
 XX
 CC The present invention relates to a new isolated dual specificity
 CC phosphatase (DUSP10) polypeptide, comprising a 665 residue amino acid
 CC sequence that is fully defined in the specification. The invention also
 CC provides a sequence encoded by a 3059 nucleotide sequence fully defined
 CC in the specification, and a sequence having at least 95 % identity to the
 CC polypeptide, or fragments or variants of DUSP-10. The invention is useful

CC for treating cancer e.g. leukaemia, colon carcinoma, lung cancer,
 CC prostate cancer, metastasis of tumour cells, neo-angiogenesis, epilepsy,
 CC stroke, neuronal degeneration syndromes, Alzheimer's disease, depression,
 CC schizophrenia, cardiac myopathies, asthma, immune disorders,
 CC inflammatory processes e.g. arthritis, bowel disease, type I diabetes,
 CC osteoporosis, diabetes and diabetes associated diseases. The molecules of
 CC the invention are also useful as vaccines for inducing immunological
 CC response in a mammal, in disease diagnosis and in assays for screening
 CC agonistic or antagonistic compounds. Other uses of the invention include
 CC identifying membrane bound or soluble receptors, as a diagnostic reagent,
 CC in chromosome localization studies, and as a valuable tool in tissue
 CC expression studies. The present sequence represents cDNA of the human
 CC dual specificity phosphatase, DUSP-10. DUSP-10 is located on chromosome
 CC 12
 XX
 SQ Sequence 3059 BP, 831 A, 761 C, 709 G, 758 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 0 Length: 3059
 Score: 472.00 Matches: 663
 Percent Similarity: 99.40% Conservative: 0
 Best Local Similarity: 99.40% Mismatches: 2
 Query Match: 70.98% Indels: 4
 DB: Gaps: 0

US-10-029-345A-109 (1-665) x AAS15768 (1-3059)

QY 1 MetAlaHisGluMetIleGlyThrGlnIleValThrGluArgLeuValAlaLeuGln 20
 Db 127 ATGGCCCATGAGATGATGTAAGTCAATTTGTTACTGAGAGGTTGGTCTGTGCGAA 186
 QY 21 SerGlyThrGluValLeuLeuIleAspSerArgProPheValGluYrAntThrSer 40
 Db 187 AGTGAACGGAAGAAAGTCTGCTAATGATGACCGGCAATTTGGAAATACATATAC 246
 QY 41 HisIleLeuGlnAlaIleAsnIleAsnCySserIlyeLeuMetIlyeArgIleGlnGln 60
 Db 247 CACATTTTGGAGCCATTATATCACTCTCCAGCTTATGAGGAAAGTTGGCAACG 306
 QY 61 AspIlyeValLeuIleThrGluLeuIleGlnHisSerAlaIlyeHisIlyeValAspIle 80
 Db 307 GACAAAGTGTATATTAACAGAGCTCATCCAGATTCAGCAACATTAAGTTGACATTGAT 366
 QY 81 CySserGlnIlyeValValIlyeValIlyeValIlyeValIlyeValIlyeValIlye 100
 Db 367 TCGAGTCAGAGGTTGATGATTCATCAAGTCCCAAGATGTCCTCTCTCTTCA 426
 QY 101 AspCyPheLeuThrValLeuLeuGlyIlyeValIlyeValIlyeValIlyeValIlye 120
 Db 427 GACTGTTTCTCACTGTAATCTTGGGTAACTGGAGAACGCTTCAACTCTGTTCACCTG 486
 QY 121 LeuAlaGlyIlyePheAlaGluPheSerArgCyPheProGlyLeuCySglGlyIlyeSer 140
 Db 487 CTGGCAGGTGGGTTTGTGATGTTCTCTGTTTCTGCGCCTGTGAAGGAAATCC 546
 QY 141 ThrIleValProThrCyValIleSerGlnProCySbleuProValAlaAsnIleGlyProThr 160
 Db 547 ACTTAGTCCCTACCTGCAATTCCTCAGCTTCTTCACTGTCGCAACATGGGCCAACCC 606
 QY 161 ArgIleLeuProAsnLeuIlyeLeuGlyCySglArgAspValIleuAsnIlyeGluIle 180
 Db 607 CGAATTTCTCCAACTTTATCTTGGCTGCCAGAGATGCTCTCAACAGAGAGCTGAT- 665
 QY 181 -GlnGlnGlnIlyeIlyeIlyeValIleuAsnIleSerTy-ThrCySProIlyeProAsp 200
 Db 666 GCAGCAGAAATGGATGGTTATGTATTAATCCAGCAA-TACCTGTCCAAAGCCGTGACT 724
 QY 200 heIleProGluSerHisPheLeuArgValProValAsnAspSerPheCySglIlyeIle 220
 Db 725 TTAATCCCGAGCTCATTTCTGCGGTGCTGTGAATGACGCTTTGTGAGAAATTT 784
 QY 220 euProThrLeuAspIlyeSerValAspPheIleGluIlyeValIlyeValIlyeValIlye 240

Best Local Similarity: 99.40% Mismatches: 2
Query Match: 70.98% Indels: 4
DB: 6 Gaps: 0

US-10-029-345a-109 (1-665) x ABN59704 (1-3104)

```
QY 1 MetAlaHisGluMetIleGlyThrGlnIleValThrGluArgLeuValAlaLeuLeuGlu 20
DB 159 ATGCCCATGATGATGGAACTCAATATTGTAAGAGGTGGTGGCTGCTGGAGAA 218
QY 21 SerGlyThrGluIleValLeuLeuIleAspSerArgProPheValGluIleThrThrSer 40
DB 219 AGTGAACCGGAAAAAGGCTGCTAATTGATAGCCGCGCATTTGTGGATTCATACATCC 278
QY 41 HisIleLeuGluAlaIleAsnIleAsnCySerIleuMetIleAsnArgLeuGlnGln 60
DB 279 CACATTTTGGAAAGCATTAATATCACTGCTCCAGCTTATGAAAGCAAGGTTGGACAG 338
QY 61 AspIleValIleuIleThrGluLeuIleGlnHisSerAlaIleValHisValAspIleAsp 80
DB 339 GACAAAGTGTAAATTACAGAGCTCATCCAGCATTCAGCAACATTAAGTTGACATTTGAT 398
QY 81 CySerGluIleValIleValIleValIleValIleValIleValIleValIleValIleVal 100
DB 399 TGAAGTCAGAGGTTGATGATTAAGATCAAAAGCTCCCAAGATGTGCTCTCTCTCTCA 458
QY 101 AspCyPheLeuThrValIleuLeuGlyIleValLeuGluIleValSerPheAsnSerValHisIleu 120
DB 459 GACTGTTTTCATCTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 518
QY 121 LeuAlaGlyGlyPheAlaGluPheSerArgCyPheProGlyLeuCyGluGlyIleValSer 140
DB 519 CTGCAAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 578
QY 141 ThrIleuValProThrCyValIleSerGlnProCyLeuProValAlaAsnIleGlyProThr 160
DB 579 ACCTAGTCCCTTACCTGATCTTCTGAGCTTCTGAGCTTCTGAGCTTCTGAGCTTCTGAG 638
QY 161 ArgIleLeuProAsnLeuIleValLeuGlyCyGlnArgAspValIleuAsnIleGlyIleu 180
DB 639 CGAATCTTCCCAATCTTATCTTGGCTGACAGAGATGCTCTCAACAGAGGCTGAT- 697
QY 181 -GlnGlnAsnGlyIleGlyIleValIleuAsnAlaSerTyr-ThrCyProIleProAsp 200
DB 698 GCAGCAGAAATGGATGGTATGTTAATGACAGAA-TACTGCTCCAAAGCTGACT 756
QY 200 IleIleProGluSerHisPheLeuArgValProValAsnAspSerPheCyGluIleIle 220
DB 757 TTATCCCGAGCTCAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 816
QY 220 euProIlePheAspIleSerValAspPheIleGluIleValAlaIleValAsnIleGlyIle 240
DB 817 TGCCTGGTGGTGAACAATCGATATTCATTAAGAAAGCAAAAGCTCCAAAGGATGG 876
QY 240 AlIleuValHisCySLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaIleTyrIle 260
DB 877 TTCTATGACACTTTTACCTGGATCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCT 936
QY 260 eIleValArgMetAspMetSerLeuAspGluAlaIleValIleValIleValIleValIleVal 280
DB 937 TGAAGGAGATGACATGCTTTAGATGAGCTTACAGATTTGGTGAAGAAAAAGACCTTA 996
QY 280 hIleSerProAsnPheAsnPheLeuGlyIleLeuLeuAspTyrGlyIleValIleValIleVal 300
DB 997 CTATATCTCCCAATCTTCTGCGCACTCTGAGCTATGAGAAAGAAATTAACA 1056
QY 300 snGlnThrGlyIleSerGlyProIleSerIleValIleValIleValIleValIleValIleVal 320
DB 1057 ACCAGACTGAGAGATCAAGGCGCAAAAGCAAACTCAAGCTGCTGCACTCGAGAACCCAA 1116
QY 320 snGluProValProAlaValIleSerGlyIleValIleValIleValIleValIleValIleVal 340
DB 1117 ATGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1176
```

```
QY 340 roCyAlaAspSerAlaThrSerGluAlaAlaGlyGlnArgProValHisProAlaSerVal 360
DB 1177 CCTGTGCCACTCTGTGTACTCAGAGCGACAGCAAAAGCCCGTGCATCCGCCAGCG 1236
QY 360 aIleProSerValProSerValGlnProSerLeuLeuGluAspSerProLeuValGlnAlaIle 380
DB 1237 TGCCACGCTGCGCCAGCGTACAGCGCTGCTGTTAGAGACAGCCGCTGTACAGCGCG 1296
QY 380 euSerGlyIleuHisIleuSerAlaAspArgLeuGluAspSerIleuValIleValIleValIleVal 400
DB 1297 TCAGTGGGTGACCCCTGCTCCGACAGCGCTGGAACAGCAATAGCTCAAGCTTCTCC 1356
QY 400 hSerLeuAspIleIleValSerValSerTyrSerAlaSerMetAlaIleSerLeuHisGlyP 420
DB 1357 TCTCTGATATCAATATGATGTTTCTATTCACCGCATGGCAGATCTTACATAGGCT 1416
QY 420 hSerSerSerGluAspAlaLeuGluTyrTyrIlePheSerThrThrLeuAspGlyThrA 440
DB 1417 TCTCTCATCAGAAAGATGCTTTGGAATATCTAACAACTTCCACTGATCGAATGGAGCA 1476
QY 440 snIleLeuCySglnPheSerProValGlnGluLeuSerGluGlnThrProGluThrSerP 460
DB 1477 ACAAGCTATGCCAGTCTCCCTGTTACAGAACTATCGAGCAGACTCCCGAAACCAAGTC 1536
QY 460 roAspIleGlyGluAlaSerIleProIleValLeuGlnThrAlaArgProSerAspSerG 480
DB 1537 CTGATTAAGAGAGAGCAGATCCCGCAAGCTGAGACCGCGAGCTTCCAGACAGCC 1596
QY 480 IserIleValArgLeuHisSerValArgThrSerSerSerGlyThrAlaGlnArgSerLeu 500
DB 1597 AGAGCAAGGATTTGATTCCTGTCAGAACCAAGCAGATGGCACCGCCAGAGGTCCCTTT 1656
QY 500 euSerProLeuHisArgSerGlySerValGluAspAsnTyrHisThrSerPheLeuPheG 520
DB 1657 TATCTCCACTGCATGCAAGTGGAGCGTGGAGAGCAATTAACACACACTTCTTTG 1716
QY 520 IleuSerThrSerGlnGlnHisIleuThrIleValSerAlaGlyIleuGlyIleuValGlyTyrH 540
DB 1717 GCTTTTCCACAGCAGCAGCAGCCTTCCAGATCTGCTGCTGCTGCTTAAAGGCTGAC 1776
QY 540 IserAspIleLeuAlaProGlnThrSerThrProSerLeuThrSerSerTyrPheA 560
DB 1777 ACTCGATATCTTGGCCCCCAGACTCTACCCCTTCCCTGACAGCAGCTGTATTTTG 1836
QY 560 IeThrGluSerSerHisPheTyrSerAlaSerAlaIleTyrGlyIleSerAlaSerTyrS 580
DB 1837 CCACAGAGCTCTCACACTTCTACTGCTGCTGAGCATCTACGAGGAGGACAGTCACTTACT 1896
QY 580 eAlaIleTyrSerCySglnPheLeuProThrCyGlyIleAspGlnValIleValIleValIleVal 600
DB 1897 CTGCTTACAGCTGACAGCCAGCTGCTCCACTTGGGAGAACCAAGTCTATTTCTGCGCAGCG 1956
QY 600 rGlnIlePheProSerAspArgAlaAspSerArgIleSerTyrHisGluGluSerProPheG 620
DB 1957 GGCAGAGCCAGATGACAGCTTGACTGCGCGGAGCTGGGATGAGAGACCCCTTTG 2016
QY 620 IuIleGlnPheIleValArgSerCySglnMetGluPheGlyIleuSerIleMetSerGluA 640
DB 2017 AAAACAGATTTAAAGCAGAAAGCTGCCAAATGGAATTTGAGAGAGCATCATGTCAAGA 2076
QY 640 snArgSerArgGluGluLeuGlyIleValIleValIleValIleValIleValIleValIleVal 660
DB 2077 ACAGGTACAGGAAAGCTGGGGAAAGGGGACAGTACGTTCCTTTCGGGACAGATGG 2136
QY 660 IuIleIleGluValSer 665
DB 2137 AATATCTGAGGTCTCC 2153
```

RESULT 8
ABK48378
ID ABK48378 strand: cDNA, 3332 BP.
XX

AC ABK48378;
XX
XX 02-JUL-2002 (first entry)
XX
XX CDNA encoding human DSP-16 alternative form protein.
DE
XX Human; dual-specificity phosphatase-3; DSP-16; autoimmune disease;
KW mitogen-activated protein kinase; MAP; Duchenne muscular dystrophy;
KW cancer; graft-versus-host disease; allergy; metabolic disease;
KW abnormal cell growth; abnormal cell proliferation; contact inhibition;
KW cell cycle abnormality; anchorage independent cell growth; apoptosis;
KW intercellular adhesion; DSP-16 modulator; chromosome 12p; gene; ss.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH 842..2395
FT CDS /tag=a
FT /product="Human dual-specificity phosphatase-3 (DSP-16)
FT alternative form protein"
FT
FT
FT
PN WO200226997-A2.
XX
XX 04-APR-2002.
XX
XX 25-SEP-2001; 2001WO-US030124.
XX
XX 26-SEP-2000; 2000US-0235487P.
XX
XX (CEPT-) CEPTYR INC.
XX
XX Lucie RM, Wei B;
XX
XX MPI; 2002-315802/35.
DR P-PSDB; AAU79159.
XX
XX New DSP-16 polypeptide, useful for identifying modulators of its
PT activity, which can be used in the treatment of disorders such as
PT Duchenne muscular dystrophy, or cancer.
XX
XX Claim 56; Fig 3; 87pp; English.
XX
XX The present invention relates to a new polypeptide, DSP-16, having a 665
CC amino acid sequence, given in the specification, or a variant having at
CC least 50 % identical residues, which retains the ability to
CC dephosphorylate an activated mitogen-activated protein (MAP) kinase. The
CC invention can be used for identifying agents which modulate DSP-16
CC activity, for modulation of a proliferative response in a cell, survival
CC of a cell, or differentiation of a cell. The cell displays contact
CC inhibition of cell growth or anchorage independent growth and may display
CC altered intercellular adhesion. The agent may modulate apoptosis, or the
CC cell cycle. The identified modulators can be used to treat Duchenne
CC muscular dystrophy, cancer, graft-versus-host disease, autoimmune
CC diseases, allergies, metabolic diseases, abnormal cell growth, abnormal
CC cell proliferation, and cell cycle abnormalities. The present nucleic
CC acid sequence is that of the human dual-specificity phosphatase-3 (DSP-
CC 16) gene located on chromosome 12p. This sequence encodes the human DSP-
CC 16 alternative form protein of the invention
XX
XX
SQ Sequence 3332 BP; 909 A; 805 C; 823 G; 795 T; 0 U; 0 Other;
Alignment Scores:
Pred. No.: 0 Length: 3332
Score: 472.00 Matches: 472
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 70.98% Indels: 0
DB: 6 Gaps: 0
US-10-029-345A-109 (1-665) x ABK48378 (1-3332)
QY 194 ThrCyProLysProAspPheIleProGluSerHisPheLeuAlaProValIAsnAsp 213
|||||

Db 977 ACCTGTCAAAGCCGATCTTATCCCGAGTCATTTCCGCGTGCCTGTGAATGAC 1036
QY 214 SerPheCyGluValIleLeuProTrieuAspLysSerValAspPheIleGluVala 233
Db 1037 AGCTTTGTGAGAAATTTTCCGCGTGTGACAAATCAGTATGATTTCATTGAGAAACA 1096
QY 234 LysAlaSerAnGlyCyValLeuValHisGlySerLeuAlaGlyIleSerArgSerAlaThr 253
Db 1097 AAAGCTCCCAATGAGTGTGTTCTAGTGCATGTTTACGTGGATTCCTCCGCTCCGAC 1156
QY 254 IleAlaIleAlaTyrIleMetLysArgMetAspMetSerLeuAspGluAlaTyrArgPhe 273
Db 1157 ATCGCTATCGCTCATCATGATGAGATGATGATGATGATGATGATGATGATGATGAT 1216
QY 274 ValLysGluLysArgProThrIleSerProAsnPheAsnPheLeuGluLeuLeuAsp 293
Db 1217 GTGAAGAGAAAAAGCCTACTATATCTCAAACTTCATTTTCGGGCACTCTGAC 1276
QY 294 TyrGluLysLysIleLysAsnGlnThrGluAlaSerGlyProLysSerLysLeuLysLeu 313
Db 1277 TATGAGAAAGATTAAGAACCAAGCTGAGCATGAGGCGCAAGACCAAACTCAAGCTG 1336
QY 314 LeuHisLeuGluLysProAsnGluProValProAlaValSerGluGlyGluLysSer 333
Db 1337 CTGCACTCGAAGAACCAATGAACCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1396
QY 334 GluThrProLeuSerProProCysAlaAspSerAlaThrSerGluAlaAlaGlyGlnArg 353
Db 1397 GAGAGCCCTCAGTCCAGCCCTGTCGACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1456
QY 354 ProValHisProAlaSerValProSerValProSerValGlnProSerLeuLeuAsp 373
Db 1457 CCCGTCATCCCGCAGCGTCCCGCAGCGTCCCGCAGCGTCCCGCAGCGTCCCGCAGCG 1516
QY 374 SerProLeuValGlnAlaLeuSerGlyLeuHisLeuSerAlaAspArgLeuLysSer 393
Db 1517 AGCCGCTGATCAAGCGCTCAGTGGCTGACCTGTCGCGAGACAGCGCTGGAACAGC 1576
QY 394 AsnLysLeuLysArgSerPheSerLeuAspIleLysSerValSerTyrSerAlaSerMet 413
Db 1577 AATAAGCTCAAGCGTCTCTCTCTGATACCAATCAGTTTCATTCAGCCAGCAGCAG 1636
QY 414 AlaAlaSerLeuHisGlyPheSerSerSerGluAspAlaLeuGluTyrTyrLysProSer 433
Db 1637 GCAGATCTTTCATGAGCTTCTCTCTCATCAGAAAGTCTTGGATATCAACAACTTCC 1696
QY 434 ThrThrLeuAspGlyThrAsnLysLeuGlyGlnPheSerProValGlnGluLeuSerGlu 453
Db 1697 ACTACTCTGATGAGGACCAACAGCTATGCCAGTCTCTCCCTGTCAGGAATATCGGAG 1756
QY 454 GlnThrProGluThrSerProAspLysGluGluAlaSerIleProLysLysLeuGlnThr 473
Db 1757 CAGACTCCGAAACCAAGTCTGATAGAGAGAGAGAGAGATCCCAAGAGCTTGCAGAC 1816
QY 474 AlaArgProSerAspSerGlnSerLysArgLeuHisSerValArgThrSerSerSerGly 493
Db 1817 GCCAGGCTTTCAGACAGCAAGCAAGCATTTGGTGTGAGAACAGCAGCAGCTGGC 1876
QY 494 ThrAlaGlnArgSerLeuLeuSerProLeuHisArgSerGlySerValGluAspAsnTyr 513
Db 1877 ACCGCCAGAGGTCCCTTTATCTCCAGTGCATCAAGAGTGGAGGTGAGACAAATTAC 1936
QY 514 HisThrSerPheLeuPheGlyLysSerThrSerGlnGlnHisLeuThrLysSerAlaGly 533
Db 1937 CACACAGCTTCTTTTGGCTTTCCACACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1996
QY 534 LeuGlyLeuLysGlyTyrHisSerAspIleLeuAlaProGlnThrSerThrProSerLeu 553
Db 1997 CTGGGCTTAAGGCTGAGCACTCGATATCTTGGCCCCCAAGACCTTACCCCTTCCCTG 2056
QY 554 ThrSerSerTyrTyrPheAlaThrGluSerSerHisPheTyrSerAlaSerAlaIleTyr 573
Db 2057 ACCACAGCTGATATTTGCCACAGAGTCTTACACTTCTACTGCTGCTGCTGCTGCTGCT 2116

QY 574 GLYGLSerAlaSerTyrSerAlaTyrSerCysSerGlnLeuProThrCysGlyAspGln 593
DB 2117 GGAGGAGATGCCAGTACTGCTGCTACAGCTGACGACGCTGCTGCGAGACCAA 2176
QY 594 ValTyrSerValArgArgGlnTyrProSerAspArgAlaAspSerArgArgSerTyrP 613
DB 2177 GCTATTCTGTGGCGAGCGGGGAGGAGCCAGAGTACGACAGCTGCTGCGCGGAGCTGG 2236
QY 614 HlaGluGluSerProPheGluTyrGlnPheLysArgArgSerCysGlnMetGluPheGly 633
DB 2237 CATGAAGAGAGCCCTTTGAAAAGCACTTTAAACGAGAGAGCTGCCAAATGGAATTTGGA 2296
QY 634 GluSerTlMetSerGluAsnArgSerArgGluGluLeuGlyValGlySerGlnSer 653
DB 2297 GAGAGCATCTGTCAGAGAACAGAGTCCGAGAAAGCTGGGAGAAAGTGGGAGTCACTCT 2356
QY 654 SerPheSerGlySerMetGluLeuIleGluValSer 665
DB 2357 AGCTTTTGGGAGCATGGAATCATTTAGAGTCTCC 2392
RESULT 9
ABK47596
ID ABK47596 standard; cDNA; 3496 BP.
XX
AC ABK47596;
DT 02-UTL-2002 (first entry)
XX
DE cDNA encoding human dual-specificity phosphatase-3 (DSP-16) protein.
XX
KW Human; dual-specificity phosphatase-3; DSP-16; autoimmune disease;
KW mitogen-activated protein kinase; MAP; Duchenne muscular dystrophy;
KW cancer; graft-versus-host disease; allergy; metabolic disease;
KW abnormal cell growth; abnormal cell proliferation; contact inhibition;
KW cell cycle abnormality; anchorage independent cell growth; apoptosis;
KW intercellular adhesion; DSP-16 modulator; chromosome 12p; gene; ss.
XX
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FH CDS 562..2559
FT /tag= a
FT /product= "Human dual-specificity phosphatase-3 (DSP-16)
FT protein"
FT
PN WO200226997-A2.
XX
XX 04-APR-2002.
PD
XX
PF 25-SEP-2001; 2001WO-US030124.
XX
XX 26-SEP-2000; 2000US-0235487P.
PR
XX
PA (CEPR-) CEPTYR INC.
XX
XX Luche RM, Wei B;
PI
XX
XX MPI; 2002-315802/35.
DR P-PSDB; AAU79156.
XX
XX
PT New DSP-16 polypeptide, useful for identifying modulators of its
PT activity, which can be used in the treatment of disorders such as
PT Duchenne muscular dystrophy, or cancer.
XX
XX
PS
XX
XX Claim 7; Fig 1; 87pp; English.
XX
XX The present invention relates to a new polypeptide, DSP-16, having a 665
XX amino acid sequence, given in the specification, or a variant having at
XX least 50 % identical residues, which retains the ability to
XX phosphorylate an activated mitogen-activated protein (MAP) kinase. The
XX invention can be used for identifying agents which modulate DSP-16
XX activity, for modulation of a proliferative response in a cell, survival

CC of a cell, or differentiation of a cell. The cell displays contact
CC inhibition of cell growth or anchorage independent growth and may display
CC altered intercellular adhesion. The agent may modulate apoptosis, or the
CC cell cycle. The identified modulators can be used to treat Duchenne
CC muscular dystrophy, cancer, graft-versus-host disease, autoimmune
CC diseases, allergies, metabolic diseases, abnormal cell growth, abnormal
CC cell proliferation, and cell cycle abnormalities. The present nucleic
CC acid sequence is that of the human dual-specificity phosphatase-3 (DSP-
CC 16) gene located on chromosome 12p. This sequence encodes the human DSP-
CC 16 protein of the invention
XX
SQ Sequence 3496 BP; 939 A; 854 C; 856 G; 847 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 0 Length: 3496
Score: 472.00 Matches: 653
Percent Similarity: 99.40% Conservative: 0
Best Local Similarity: 99.40% Mismatches: 2
Query Match: 70.98% Indels: 4
DB: 6 Gaps: 0
US-10-029-345A-109 (1-665) x ABK47596 (1-3496)
QY 1 MetAlaHisGluMetIleGlyThrGlnIleValThrGluArgLeuValAlaLeuLeuGlu 20
DB 562 ATGCCCATGAGATGATTGGAATCTCAATTTGATTCAGAGAGTGTGGCTCTGCTGGA 621
QY 21 SerGlyThrGluLysValLeuLeuLeuIleAspSerArgProPheValGluTyrAsnThrSer 40
DB 622 ACTGGAACCGAAGAAAGTGCTGCTTAATGTATGATCCGGCCATTGTGGAAATACATATAC 681
QY 41 HisIleLeuGluAlaIleAsnIleAsnIleAsnIleAsnIleAsnIleAsnIleAsnIle 60
DB 682 CACATTTTGGAAGCCATTAAATCAACGCTCCAACTTATAGAGGAGTTGCAACAG 741
QY 61 AspLysValLeuIleThrGluLeuLeuIleGlnHisSerAlaLysIleValAspIleAsp 80
DB 742 GACAAAGTCTTAATTAACAGAGCTCATCCAGCTTCAGGCCAAACATTAAGTTGACATTGAT 801
QY 81 CysSerGlnLysValValValValTyrAspGlnSerSerGlnAspValAlaSerLeuSerSer 100
DB 802 TCAGTACGAAAGATTGATGATTAACGATCAAACTCCCAAGATGTTGCTCTCTCTTCA 861
QY 101 AspCysPheLeuThrValLeuLeuGlyLysLeuGluLysSerPheAsnSerValHisLeu 120
DB 862 GACTGTTTCTCAGTACTTCTGGGTAAACGAGAAAGCTTCAACTCTGTCACTG 921
QY 121 LeuAlaGlyGlyPheAlaGluPheSerArgCysPheProGlyLeuCysGluGlyLysSer 140
DB 922 CTTCAGAGTGGGTTTGTCTGAGTTCTCTCTCTTCTCTCTCTCTCTCTCTCTCTCTCA 981
QY 141 ThrLeuValProThrCysIleSerGlnProCysLeuProValAlaAsnIleGlyProThr 160
DB 982 ACTCTAGTCCCTACCTGATTTCTCAGCTTCTTACCTTACCTGCTTACCTGCTTACCT 1041
QY 161 ArgIleLeuProAsnLeuTyrLeuGlyCysGlnArgAspValLeuAsnLysGluLeuIle 180
DB 1042 CGAATTCCTCCAAATCTTATCTTGTGCTGCCAGCAGATGCTCTCAACAGAGCTGAT- 1100
QY 181 -GlnGlnAsnGlyIleGlyTyrValLeuAsnAlaSerTyr-ThrCysProLysProAsp 200
DB 1101 GCAGCAGATGGAGATTGTTATGTTTAAATCCAGCAA-TTCCTGCTCAAGAGCTGACT 1159
QY 200 IleIleProGluSerHisPheLeuArgValProValAsnAspSerPheCysGluLysIle 220
DB 1160 TTATCCCGAGTCTCATTTCTCGCGTGGCTGTGAATGACGCTTTGTGAAATTT 1219
QY 220 euProTTrpLeuAspLysSerValAspPheIleGluLysAlaLysAlaSerAsnGlyCysV 240
DB 1220 TCCCCTGTGTGACAAATCAGATGATTCATTGAAAGCAAAAGCAAGCTTCAATGATG 1279
QY 240 alleuValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIleM 260

Db 1280 TTCTAGTGCATGTTTAGTGTGGATCTCCGCTCCGACACCATGGCTATGCGCTACATCA 1339
 QY 260 etlysaagmetaspmetserleuaspglualatyargphevallylsglylysaargprot 280
 Db 1340 TGAAGAGGATGACATGCTTTAGATGAACTTACGATTTTGAAAGAAAAAGACCTTA 1399
 QY 280 hrileserproanpheasnphelenglvglinleuaspkyrfglylylysailelysa 300
 Db 1400 CTATATCTCCAAACTTCAATTTTCTGGGCCCACTCTCGACTATGAGAAAGAAATTAAGA 1459
 QY 300 snglnthrghlalaserglylprolysserlylseuysleuenuhlsleuglylysa 320
 Db 1460 ACCAGACTGGAGCATGAGGSCCAAGAGCAACTCAAGCTGCTGACCTGGAGAAACCA 1519
 QY 320 snglnuprovalproalavalserglvglyglinlyssergluthrproleuserprop 340
 Db 1520 ATGAACCTGTCCTGCTGCTCTCAGAGGCTGAGCAAGAAAGCGAGAGCCCTTCAGTCAC 1579
 QY 340 rocyaslaaspserrallathrserglualalaglvglnaargprovalhlsproalaser 360
 Db 1580 CCTGTGCCACCTCTGCTACTCTCAGAGGCAAGCAAGAAAGCCGCTGCAATCCGCGCACG 1639
 QY 360 alproservalproservalginproserleuenglunaspserproleuvalglinalal 380
 Db 1640 TGCCCAAGCTGCGCCAGCGTGCAGCCGTCGCTGTAGAGAGACGCCGCTGTATACAGGCC 1699
 QY 380 euseerlyleuuhlsleuseralaaasparglenglunasperenlylylylysaargserp 400
 Db 1700 TCAGTGGGCTGCACCTGCTCCGAGACAGCTGGAACAGCAATAGCTCAAGCCGTCCT 1759
 QY 400 hsserleuaspillelysserlaserlyserlaserlaserlaserlaserlaserlaser 420
 Db 1760 TCTCTGTGATATCAATCAAGTTTCAATATTCAGCCAGCATGGCAGATCTTACATGCT 1819
 QY 420 hsserlaserlaserlaserlaserlaserlaserlaserlaserlaserlaserlaser 440
 Db 1820 TCTCCCATCAGAAAGATGCTTGGATATCTACAAACCTTCCACTGCTGATGGAGCA 1879
 QY 440 snlylseuysglinpbeserprovalglnleuaserluginthrprogluthrserp 460
 Db 1880 ACAAGCTATGCAAGTCTCCCTGTCAGAACTATCGAGAGACATCCCGAAACAGAGC 1939
 QY 460 roasplysglunlaserlaserlaserlaserlaserlaserlaserlaserlaserlaser 480
 Db 1940 CTGATAGAGAGGAGGACCAATCCCAAGAGCTGAGAGCCGCGCTTACAGACGCC 1999
 QY 480 lnserylsargleuuhlssevalargthrserlaserlaserlaserlaserlaserlaser 500
 Db 2000 AGAGCAAGGATGATGCTGCTCAGAACCAAGAGAGTGGACCGCCAGAGCTCCCTTT 2059
 QY 500 euseerproleuuhlsargsergllyservalglunaspasnthrsrphelaupheg 520
 Db 2060 TATCTTCATCTGATGAGAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2119
 QY 520 lyseuserthrserglnglnhlsleuthrlysserlaserlaserlaserlaserlaser 540
 Db 2120 GCCTTTCCACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 2179
 QY 540 isseraspilleleualaproglinthrserthrproserleuuthrserlaserlaser 560
 Db 2180 ACTCGGAATATCTTGAGCCGCGAGACCTTACCCCTTCCGACGAGCTGGATTTTGG 2239
 QY 560 lathgluserserhlsphetyrserlaserlaserlaserlaserlaserlaserlaser 580
 Db 2240 CCACGAGTCTCTCACTTCTACTCTGCTCAGCCATCTACGAGGAGGAGGAGGAGGAGGAG 2299
 QY 580 ealalytsercysserglneuprothrlysglyaspglinalalytserlaserlaser 600
 Db 2300 CTGCTTACAGCTGAGCCAGCGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2359
 QY 600 rglnglylproseraspargalaaasparargargserlaserlaserlaserlaserlaser 620
 Db 2360 GGCAGAGCCAAAGTACAGAGCTGACTGCTGCGGCGAGCTGCGATGAAAGAGAGCCCTTTG 2419

QY 620 lulyeGlnpheylsargargsercyssglmetglupheglvglnserlaserlaserlaser 640
 Db 2420 AAAACAGATTTAAAGCAGAGAGCTGCCAAATGGAATTTGGAGAGAGCATGTCAGAGA 2479
 QY 640 snargserarglglngluegllylysaValgylserglinserserpheserlaserlaser 660
 Db 2480 ACAGSTCACGGGAGAGCTGGGAGAAAGTGGCAGTCACTTCTTTCGGGACAGCATGG 2539
 QY 660 lullellegluvalser 665
 Db 2540 AAATCATTTAGAGTCTCC 2556
 Db 2540 AAATCATTTAGAGTCTCC 2556
 RESULT 10
 AAS14639
 ID AAS14639 standard, cDNA: 3544 BP.
 XX
 AC AAS14639;
 DT 18-DEC-2001 (first entry)
 XX
 DE Human cDNA encoding dual specificity phosphatase 21117.
 KW Human; ss: dual specificity phosphatase 21117; hepatotropic; cytostatic;
 KW hematopoietic disorder; autoimmune disorder; diabetes mellitus;
 KW rheumatoid arthritis; multiple sclerosis; Crohn's disease;
 KW liver disorder; erythroid associated disorder; haemolytic anaemia;
 KW cellular proliferative; differentiative disorder; leukaemia;
 KW acute myeloid leukaemia; carcinoma; sarcoma; metastatic cancer.
 XX
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 589..2586
 FT /*tag= a
 FT /product= "Dual specificity phosphatase 21117"
 XX
 PN WO200173059-A2.
 XX
 PD 04-OCT-2001.
 XX
 PF 23-MAR-2001; 2001WO-US009477.
 XX
 PR 24-MAR-2000; 2000US-0191858P.
 XX
 PA (MILL-) MILLENNIUM PHARM INC.
 XX
 PI Meyers RA;
 XX
 DR WPI: 2001-611635/70.
 DR P-PSDB; AAU09016.
 PT New human dual specificity polypeptides and nucleic acids for diagnosis
 PT of disease and treatment of e.g. liver disorders.
 XX
 PS Claim 1; Fig 1; 143bp; English.
 XX
 CC The invention relates to two novel human dual specificity phosphatases
 CC designated 21117 and 38692, the nucleic acids encoding them (including
 CC fragments, allelic variants, their complements or nucleic acids that
 CC hybridise to them) and antibodies raised against the proteins. The
 CC antibody is useful for detecting the presence of the polypeptide, and the
 CC nucleic acid fragments are useful for detecting the presence of the
 CC nucleic acid, for diagnosing disease. Agents (e.g. the antibody and
 CC antisense sequences) are useful for modulating the activity or expression
 CC of the polypeptide/nucleic acid, useful for the treatment of e.g.
 CC hematopoietic disorders, autoimmune disorders (e.g. diabetes mellitus,
 CC rheumatoid arthritis, multiple sclerosis, Crohn's disease and many others
 CC listed in the specification) liver disorders, erythroid associated
 CC disorders (e.g. haemolytic anaemia) cellular proliferative or
 CC differentiative disorders, leukaemias (e.g. acute myeloid leukaemia),
 CC carcinomas, sarcomas, metastatic cancers. Agents modulating 21117 and
 CC 38692 are also useful for modulating the proliferation, survival,

CC migration or differentiation of a 38692 or 21117-expressing cell. The
CC polypeptide and nucleic acids are useful for identifying modulating
CC agents. The present sequence encodes the dual specificity phosphatase
CC 21117

XX Sequence 3544 BP; 1000 A; 840 C; 882 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0	Length:	3544
Score:	472.00	Matches:	663
Percent Similarity:	99.40%	Conservative:	0
Best Local Similarity:	99.40%	Mismatches:	2
Query Match:	70.98%	Indels:	4
DB:	5	Gaps:	0

US-10-029-345A-109 (1-665) x AAS14639 (1-3544)

```
QY 1 MetLAHIGLUMETILEGLYTHRGINLEVALTHRGUARGLVVALALEULEUGLU 20
DB ATGGCCCATGAGATGATGGAACTCAATTGTTACTGAGAGGTGGTGGCTGCTGGA 648
QY 21 SerGIYThrGIuLYValLEULEUlleaSPSerArgProPheValGLuTYrSenthS 40
DB AGTGAACCGAAAGCTGCTGCTAAATTGATAGCCGGCCCAATTGTGGAATACATACC 708
QY 41 HIEILEUGLUALALEANILEANILEANCYSESRILEUUMETLYEARGHGLENGIN 60
DB CACATTTTGAAAGCCATTATATCACTGCTCCAACTTATGAAAGGATGGCAACG 768
QY 61 AEPLYValLEUlleThrGIuLEUlleGINHISERAlaLYSHISLYSValASPleAP 80
DB GACAAAGTGTATTCACAGGCTCATCCAGCATTCAGCAACATAGATTGACATTGAT 828
QY 769 GACAAAGTGTATTCACAGGCTCATCCAGCATTCAGCAACATAGATTGACATTGAT 828
DB 81 CYSESRGINLYValValIYrAPGINSESRSESRGINAPYAlaSERleuSER 100
DB 829 TGCAGTCAGAGGTTGATTACGATCAAGCTCCAGATGTTGCTCTCTCTTCA 888
QY 101 AEPQYPhELEUThrValLEULEUGLYrSLEUGLULYSESRPheSMBERValHISLEU 120
DB 889 GACTGTTTCTCATCTACTCTGGGTAAACGTGGAAGAGCTTCAACTCTGTCACTG 948
QY 121 LeuAlaGLYGLYPhElaGLUPheSArgCYSPheProGLYleuCYSGluGLYYSER 140
DB 949 CTTCGACGGTGGTTCCTGAGTTCTCTGTTGTTTCCCTGGCTCTGTGAAGAAATCC 1008
QY 141 ThrLEUValProThrCYsIIeSERGINProCYsLEUProValAlaANILeGLYProThr 160
DB 1009 ACTCTAGTCCCTACCTGCATTTCTCAGCCTTGCTTACTCTGTTGCCAACTTGGGCCAAC 1068
QY 161 ArgILEUProANleUThrLEUGLYCYSGINArgAPYAlaLEUanLYSGILEUlle 180
DB 1069 CGAATTCCTCCCAATCTTATCTGTGCTGCACGAGATGCTCTCAACAGAGCTGAG- 1127
QY 181 -GINGINaNGLYILEGLYTYrValLEUanlaSERTYr-ThrCYSProlYSProASP 200
DB 1128 GACGCGAATGGATTTGTTATGTGTTAAATGCCACAA-TACCTGTCCAAAGCTTGACT 1186
QY 200 heIIeProGluSERHISpHeLEUArgValProValaMBaSPSerPheCYSGIuLYSLEU 220
DB 1187 TTATCCCGAGTCTCATTTCTCGTGTGCTGTGATGACAGCTTTGTGAGAAATTT 1246
QY 220 euProTrPLEUaPlySSErValaSPheIIeGLULYsAlaLYsAlaSERanGLYCYAV 240
DB 1247 TGGCGGTGGTGGCAATCATGATGATTTCAATTGAGAAAGCAAAAGCTCCATGTGATGG 1306
QY 240 alLeuValHISCYsLEUAlaGLYILESRArgSERAlaThrIIeAlaIIeAlaTYrIIe 260
DB 1307 TTCTTAATGACCTTTTAAGCTGGAGATCTCCGCTCCGACCAATGCTTAATGCTTAATCA 1366
QY 260 eLYsArgHISpHeLEUaPlySSErLEUaPlyAlaTYrArgPheValLYSGIuLYSArgPro 280
DB 1367 TGAAGAGATGACATGCTTTTATGATGAGCTTACAGATTTGTGAAAGAAAGAAAGCACTA 1426
```

```
QY 280 hrlleSErProANpHeANpHeLEUGLYGINLEUleuAPTYrGIuLYSlySIIeLYa 300
DB 1427 CTAATCTCCAAACTTCAATTTTCTGGGCCAACTCTGGACCTATGAGAAAGATTAA 1486
QY 300 enGINThrGIYAlaSERGIYProLYSSErLYSLEUlySLEUleuHISLEUGIuLYSPRO 320
DB 1487 ACCAGACTGAGACATCAGGCGCAAGAGCAAACTCAAGCTGTGACCTGGAGAAAGCCAA 1546
QY 320 enGINProValProAlaValaSERGIUGLYGINLYSSErGIuThrProleuSERPro 340
DB 1547 ATGAACCTGTCTCTCTGTCTGAGAGGGTGGACAAAAAGCAAGAGCCCTCAGTCCAC 1606
QY 340 roCYAlaAPoSERAlaThrSERGIuAlaAGLYGINArgProValHISProAlaSERV 360
DB 1607 CCTGTGCCACTCTGCTACTACCTAGAGGACGAGCAAAAGCCCTGTGCTCCGCGACGG 1666
QY 360 alProSErValProSErValGINProSErLEUleUGIuAPSErProLEUValGINAla 380
DB 1667 TCCCGAGGCTGCCAGCGGTGAGCCGTGCTGTAGAGACAGCCGCTGTATACAGGCC 1726
QY 380 euSERGIYleuHISleuSERAlaAPArgLEUGIuAPSErANLYSleuLYSArgSERP 400
DB 1727 TCAGTGGGCTGCACCTGTCCCGACAGCGCTGGAAGACGCAATAGCTCAAGCTTCT 1786
QY 400 heSErLEUaSPILeLYSSErValaSERTYrSErAlaSERMerAlaIaSERleuHISGLY 420
DB 1787 TCTCTGATATCAATCATGTTTCATATTCAGCCAGATGGAGATCTTACATGGCT 1846
QY 420 heSErSErSErGIuAPAlaLEUGIuTYrTYrLYSProSErThrThrLEUaPGLYThr 440
DB 1847 TCTCTCATCAAGAGTGTGTTGAAATCTACAAACCTTCACATCTGTGATGGAGCA 1906
QY 440 enLYSleuCYSGINPheSErProValGINGLYleuSERGIUGINThrProGIuThrSERP 460
DB 1907 ACNAGCTATGCGAGTCTCCCTGTTCAGGAATTCGAGAGAGACTCCCGAAACCAAGTC 1966
QY 460 roAPLYSGLINuLaseRIleProLYSleuGINThrAlaArgProSARaPSErG 480
DB 1967 CTGATTAAGAGAGAGCAGACATCCCAAGAACTGCAGACCCGAGCTTTACAGACGCC 2026
QY 480 InSErLYsArgLEUHisSErValaArgThrSErSErSErGIYThrAlaGINArgSERLEU 500
DB 2027 AAGAGAGAGATGATTCGATCGACAGACAGAGAGCTGGACCCGCGAGAGTCCCTTT 2086
QY 500 euSErProleuHISArgSERGIYSErValGIuAPaNTYrHISrSErPheLEUphG 520
DB 2087 TATCTCCACTGCATGGAAGTGGAGCTGGAGAGCAATTAACAACACCTTCTTTTGG 2146
QY 520 LYleuSErThrSErGINGINHISleuThrLYSSErAlaGLYleuGLYleuLYSGIYTrH 540
DB 2147 GCCTTTCCACGACGACGACGACCTTACGAAATCTGTGGCTTGAAGGCTGGC 2206
QY 540 IASerASpIIeLEUAlaProGINThrSErThrProSErleuThrSErSErTPYrPheA 560
DB 2207 ACTCGGATATCTGGCCCCCGACCTTCACTCCCTTCCCTGACAGAGCTGGATTTTG 2266
QY 560 IATHrGluSErSErHISpHeTYrSErAlaIaSERAlaIIeTYrGIYsERAlaSERTYrS 580
DB 2267 CCACAGAGCTCCCACTTCTACTGCTTACGACATCAAGAGGAGTGCACAGTTACT 2326
QY 580 eRIaTYrSErCYsSErGINleuProThrCYsGLYAPGINValTYrSErValaArgArg 600
DB 2327 CTGCTTACAGCTGACGACGCTGCCACTTGGGAGAACCAATCTTATTTCTGTGCGAGGC 2386
QY 600 rGINLYSProSErAPArgAlaAPSErArgArgSERTPHISGLIuGluSErProPhG 620
DB 2387 GGCAGAGCAAGTACAGAGCTGACTCCGCGGAGCTGAGCATGAAGAGAGCCCTTTG 2446
QY 620 IuLYSGINPheLYsArgArgSErCYSGINMetGIuPhEGIYGIuSERIIeMetSERGIuA 640
DB 2447 AAAACAGATTTAAAGCAGAGCTCCCAATGGAATTTGAGAGAGCATCATGTGAGAGA 2506
QY 640 enArgSErArgGIuLEUGLYrSValGIYSErGINSErSErPheSErGIYSErMetG 660
```



```

QY 360 aLProSerValPProSerValGlnProSerLeuLeuGlnuAAspSerProLeuValGlnAlaL 380
DB 1667 TGGCCAGCGTGGCCAGCGTGCACCCGTCGTGTAAGAGGACACCCGCTGGTACAGGCGC 1726
QY 380 euSerGlyLeuH1LeuSerValAAspArgLeuGlnuAAspSerAenLysLeuLysArgSerP 400
DB 1727 TCAGTGGGCTGCACCTGCTCCGAGACAGGCTGGAACACAGCATTAAAGCTCAAGCCTTCCT 1786
QY 400 heSerLeuAAspL1LeuSerValSerTyrSerAlaSerMetAlaA1AserLeuH1GlyP 420
DB 1787 TCCTCTGATATCAATCAATCATGTTTCATATTCACCCGACATGCGACATCCTTACATGGCT 1846
QY 420 heSerSerSerGlnuAAspAlaLeuGlnuTyrTyrLysProSerThrThrLeuAAspGlyThra 440
DB 1847 TCTCTCATCATGAAGATGCTTGAAATCTACAAACCTTCCACTACTCTGGATGGGACCA 1906
QY 440 enLysLeuCySGlnPheSerProValGlnGlnuSerGlnuGlnThrProGluThSerP 460
DB 1907 ACAAGCTATGCGAGTCTCCCTGTTCAAGAACTATCGAGAGAGACTCCGAAACAGTC 1966
QY 460 roAAspLysGlnuAAspL1AserT1LeProLysLysLeuGlnThraL1AAspProSerAAspSerG 480
DB 1967 CTGATTAAGGAGGAGAGCAGCATCCCAAGAAAGCTGCAAGCCGACGCTTCAGACAGCC 2026
QY 480 InSerLysArgLeuH1AserValArgThrSerSerSerGlyThraL1ArgLysSerLeuL 500
DB 2027 AGAGCAAGGATGTCATTCGCTAGAAACCAAGAGCTGACAGCCGACGAGGCTCCCTT 2086
QY 500 euSerProLeuH1AAspSerGlySerValGlnuAAspAntTyrH1SerThrPheLeuPheG 520
DB 2087 TATCTCATCTGATCGAAGTGGAGGAGCTGGAGAGCAATTACCAACAGCTTCCTTTTG 2146
QY 520 LysSerThrSerGlnGlnH1LeuThrLysSerAlaGlyLeuGlyLeuLysGlyTyrP 540
DB 2147 GCCTTCCACAGCAGCAGCAGCACTCAAGAGTCTGCGCTGAGCTTAAAGGCTGGC 2206
QY 540 iSerSerP1LeuAlaProGlnThrSerThrProSerLeuThrSerSerTyrPheA 560
DB 2207 ACTGGATATCTGGCCCCCAGACCTTAACCTCCCTCCGACAGCAGCTGATATTTTG 2266
QY 560 lAthrGlnSerSerH1SphenTyrSerAlaSerAlaL1L1TyrGlyGlySerAlaSerTyrS 580
DB 2267 CCACAGAGCTCTCACCTTCTACTCTGCTCAGCCATCTACGAGGAGCAGGCTTACT 2326
QY 580 exAlaTyrSerCySGlnLeuProThrCySGlyAAspGlnValTyrSerValArgArgA 600
DB 2327 CTGCTTACAGCTGCACAGCAGCTCCCACTTGCAGAAACCAAGCTATTTCTGTCGAGGC 2386
QY 600 rGlnLysProSerAAspArgAlaAAspSerArgArgSerTyrPheLysGlnuGlnuSerProPheG 620
DB 2387 GGCAGAGCCAACTGACAGAGCTGACTCGCGGAGGAGCTGGCAATGAAAGAGGCCCTTTG 2446
QY 620 lLysGlnPheLysArgArgSerCySGlnMetGlnuPheGlyGlnuSerL1MetSerGlyA 640
DB 2447 AAAAGAGCTTTAAACGCAAGAGCTGCCAAATGGAATTTGGAGAGAGCATCATTCAGAA 2506
QY 640 enArgSerArgGlnuGlnuGlyLysValGlySerGlnuSerSerPheSerGlySerMetG 660
DB 2507 ACAGGTCACGGGAAGAGCTGGGAAAGTGGGAGTCAAGTCACTTTTCCGGCAGCATGG 2566
QY 660 lL1L1L1L1L1L1L1L1L1L1L1L1L1L1L1L1L1L1L1L1L1L1L1L1L1L1L1 665
DB 2567 AAATCATATGAGGTCTCC 2583

```

```

XX XX
KW Human; protein phosphatase 7; PP7; immune system disorder; AIDS;
KW acquired immunodeficiency syndrome; thymic hypoplasia; anemia;
KW Crohn's disease; neurological disorder; epilepsy; Huntington's disease;
KW dementia; Parkinson's disease; developmental disorder; Down's syndrome;
KW cell proliferative disorder; cancer; adenocarcinoma; leukemia; lymphoma;
KW melanoma; myeloma sarcoma; gene; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT CDS 538..2535
XX FT /tag=a
XX FT /product="Protein_phosphatase_7_(PP7)"
XX PN WO200210363-A2.
XX PD 07-FEB-2002.
XX PF 26-JUL-2001; 2001WO-US023716.
XX PP 28-JUL-2000; 2000US-0221679P.
XX PR 03-AUG-2000; 2000US-0223272P.
XX PR 10-AUG-2000; 2000US-0224309P.
XX PR 18-AUG-2000; 2000US-0226728P.
XX PR 30-AUG-2000; 2000US-0229254P.
XX PR 08-SEP-2000; 2000US-0231366P.
XX PA (INCYTE GENOMICS INC.
XX PI Tang YT, Elliott VS, Rankumar J, Yao MG, Burford N, Wang YE;
XX PI Stewer EA, Gandhi AR, Patterson C, Lee EA, Hatalla AA, Lu DM;
XX PI Trilouley CM, Griffin JA, Baughn MR, Yue H, Warren BA, Nguyen DB;
XX PI Walla NK, Kearney L;
XX DR WPI; 2002-188735/24.
XX DR P-PDB; AAU75789.
XX PT New protein phosphatases, useful for diagnosing, treating or preventing
XX PT immune system disorders (e.g. Crohn's disease); neurological disorders
XX PT (e.g. Parkinson's disease), or cell proliferative disorders (e.g.
XX PT cancers).
XX PS Claim 5; Page 114-115; 117pp; English.
XX CC The present invention relates to a new polypeptide, a naturally occurring
XX CC amino acid sequence at least 95 % identical to it, a biologically active
XX CC fragment of it or an immunogenic fragment of it. The polypeptides,
XX CC polynucleotides, agonists and antagonists are useful for diagnosing,
XX CC treating or preventing disorders associated with aberrant expression of
XX CC protein phosphatases (PP), particularly immune system disorders e.g.
XX CC acquired immunodeficiency syndrome (AIDS), thymic hypoplasia, anemia,
XX CC asthma or Crohn's disease, neurological disorders e.g. epilepsy,
XX CC Huntington's disease, dementia or Parkinson's disease, developmental
XX CC disorders e.g. Down's syndrome, or cell proliferative disorders e.g.
XX CC cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma
XX CC or sarcoma. The present nucleic acid sequence encodes human protein
XX CC phosphatase 7 (PP7) which is one of several human protein phosphatases
XX CC (AAU75783-AAU75792) of the invention
XX SQ Sequence 3766 BP; 1011 A; 903 C; 918 G; 934 T; 0 U; 0 Other;

```

Alignment Scores:

```

Pred. No.: 0 Length: 3766
Score: 472.00 Matches: 663
Percent Similarity: 99.40% Conservative: 0
Best Local Similarity: 98.40% Mismatches: 2
Query Match: 70.98% Indels: 4
DB: 6 Gaps: 0

```

US-10-029-345A-109 (1-665) x ABK14474 (1-3766)

QY 1 MetAlaH1eGlnuMetL1eGlyThrGlnL1eVal1ThrGlnuArgLeuValAlaLeuLeuGlnu 20

Db	538	ATGCCCATGAGATGATTGGAACTCAATATTGTTACTAGAGGTTGGTGGCTGCTGGAA	597
Qy	21	SeGClYthrglYluySValleuLeuLiLeaSPSerArpProPheValGluYrYantrSer	40
Db	598	AGTGAACGGAAAAAGCTGTCTAATTGATAGCGGCACTTGTGGAAATCAATACATCC	657
Qy	41	HisIleuGluAaIleAenIleAenCySerLysIleuMetLysAArgLeuGlnGln	60
Db	658	CACATTTTGGAAAGCATTAATATCACTGCTCCAAAGCTTATGAAGGAGGTGGCAACG	717
Qy	61	AspLysValleuIlethrgluleuIleGlnHisSerAlaLysHisLysValAspIleAap	80
Db	718	GACAAAGTGTAAATTAACAGAGCTCATCCAGCATTCAGCCGAAACATTAAGTTGACATTGAT	777
Qy	81	CysSerGlnLysValValAlaYrYrAspGlnSerSerGlnAspValAspIleuSerSer	100
Db	778	TGCAGTCAAGAAAGTGTAGTTACGATCAAAAGCTCCCAAGATGTGCTCTCTCTTCA	837
Qy	101	AspCysPheLeuThrValleuLeuGlyLysLeuGluLysSerPheAenSerValHisLeu	120
Db	838	GACTGTTTCTCACTGACTTCTGGGTAACTGGAAGAGCTTCAACTCTGTTCACTCG	897
Qy	121	LeuAlaGlyGlyPheAlaGluPheSerArgCysPheProGlyLeuCysGlnGlyLysSer	140
Db	898	CTTGACAGTGGGTTTCTGCTGAGTCTCGTGTGTTTCCCTGGCCTGTGTGAAGAAATCC	957
Qy	141	ThrLeuValProThrCysAlleSerGlnProCysLeuProValAlaAsnIleGlyProThr	160
Db	958	ACTCTAGTCCCTACCGACATTTCTCAGCCTTGCTGCTTGCCTGCAAACTTGGGCCAAC	1017
Qy	161	ArgIleLeuProAenLeuTYrLeuGlyCysGlnArgAspValleuAsnLysGlnLeuIle	180
Db	1018	CGAATTTCTCCCAATCTTTATCTTGCTGCCAGCGAGATGCTCTCAACAAGAGCTGAT-	1076
Qy	181	-GlnGlnAsnGlyIleGlyTYrValleuAsnAlaSerTYr-ThrCysProLysProAsp	200
Db	1077	GCAGCGAAATGGGATTGGTTATGTGTTAAATGCGACAGAA-TACCTGTCCAAAGCCTGACT	1135
Qy	200	heIleProGlnSerHisPheLeuArgValProValAsnAspSerPheCysGlnLysIleL	220
Db	1136	TATATCCCCAGATCTCATTTCTGCGGTGCTGTGATGACAGACTTGTGTGAAATTT	1195
Qy	220	eupProTyrLeuAspLysSerValAspPheIleGluLysAlaLysAlaSerAsnGlyCysV	240
Db	1196	TGCCGTGGTTGGACAAATCAGTATGATTTCATTGAGAAACAAAGCCTCCATGATGTG	1255
Qy	240	AlleuValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTYrIleW	260
Db	1256	TTCCTAGTGAAGCTGTAGCTGGGATCTCCGCTCCGCCACCATCGCTATCGCTACATCA	1315
Qy	260	eLysArgMetAspMetSerLeuAspGluAlaTYrArgPheValLysGlnLysAsnProT	280
Db	1316	TGAAGAGATGACATGCTTTTATGATGAAGCTTACGATTGTGAAAGAAAAAACCTTA	1375
Qy	280	hIleSerProAenPheAenPheLeuGlnLeuLeuAspTYrGlnLysLysIleLysA	300
Db	1376	CTATATCTCCAAACTTCAATTTTCTGGGCCAACTCTGGAATATGAGAAAGATTAGA	1435
Qy	300	snGlnThrGlyAlaSerGlyProLysSerLysLeuLysLeuLysIleuGlnLysProA	320
Db	1436	ACCAACACTGAGCAATCAGGCGCAAAAGACAAACTCAAGCTGTGCACTGGAGAAAGCAA	1495
Qy	320	snGlnProValProAlaValSerGlnGlyGlnLysSerGlnLysProLysSerProP	340
Db	1496	ATGAACCTGCTCTCTGCTCTCAAGAGGTGACACAAAAAGCAGAGCGCCCTCAGTCCAC	1555
Qy	340	roCysAlaAspSerAlaThrSerGluAlaAlaGlyGlnArgProValHisProAlaSerV	360
Db	1556	CCTGTGCCGACTGCTGCTACCTCAGAGGACGACGACAAAGGCCCTGTGATCCCGCCACG	1615
Qy	360	AlProSerValProSerValGlnProSerLeuLeuGluAspSerProLeuValGlnAlaL	380
Db	1616	TCGCCAGCGTGCACGCTGACAGCCGTGCTGTTAGAGACAGCCCCGCTGTACAGGCC	1675
Qy	380	eusSerGlyLeuHisLeuSerAlaAspArgLeuGluAspSerAsnLysLysArgSerP	400
Db	1676	TCAGTGGCTGCACTGTCCCGACAGCAGCTGAAAGACAGAAATAGCTCAAGCTTCT	1735
Qy	400	heserLeuAspIleLysSerValSerTYrSerAlaSerMetAlaAlaSerLeuHisGlyP	420
Db	1736	TCTCTGTGATATCAAAATCAGTTTCATATTCATAGCCAGCATGGACATCTTACATGGCT	1795
Qy	420	heserSerSerGlnAspAlaLeuGlnGlyTYrLysProSerThrThrLeuAspGlyThra	440
Db	1796	TCTCTTCATCAGAAATCTTCTTGAAATCTCAAACTTCCACTACTCTGTGATGGACCA	1855
Qy	440	snLysLeuCysGlnPheSerProValGlnLysLeuSerGlnGlnThrProGlnThrSerP	460
Db	1856	ACAAGCTATGCGCAGTTCTCCCTGTTCAGAACTATCGACAGACACTCCGAAACAGTC	1915
Qy	460	roAspLysGlnGluAlaSerIleProLysLysLeuGlnThrAlaArgProSerAspSerG	480
Db	1916	CTGATTAAGAGAAAGCCAGCATCCCAAGAACTGACAGACCGCCAGGCTTTCAGACAGCC	1975
Qy	480	InserLysArgLeuHisSerValArgThrSerSerSerGlyThrAlaGlnArgSerLeuL	500
Db	1976	AGAGCAAGCATTCGATTCGATCAGAACACAGACAGACAGCCGACAGAGTCCCTTT	2035
Qy	500	eusSerProLeuHisArgSerGlySerValGluAspAsnTYrHisThrSerPheLeuPheG	520
Db	2036	TATCTCCACTGCATCGAAGTGGAGCGTGGAGGACAAATTACACACAGCTTCTTTTCG	2095
Qy	520	LysSerSerThrSerGlnGlnHisLeuThrLysSerAlaGlyLeuGlyLysLeuLysGlyT	540
Db	2096	GCTTTTCCACAGCAGAGACCACTCAGAACTGTGCTGGCTGGGCTTAAAGGCTGGGC	2155
Qy	540	IsSerAspIleLeuAlaProGlnThrSerThrProSerLeuThrSerSerTyrPheA	560
Db	2156	ACTCGAATATCTTGCCCCCAGACCTTACCTTCCCTGACACAGCAGCTGTGATTTTG	2215
Qy	560	lathrgluserSerHisPheTYrSerAlaSerAlaIleTYrGlyGlySerAlaSerTYrs	580
Db	2216	CCACAGAGTCTTCACACTTCTACTCTGCTCAGACCATCTCAGGAGGACGACGAGTTACT	2275
Qy	580	erAlaTYrSerCysSerGlnLeuProThrCysGlnAspGlnValTYrSerValArgArgA	600
Db	2276	CTGCTTCAAGCTGACGACCACTGCCACTTGCGGAGCCAAAGTCTATTTGTGCGCAGGC	2335
Qy	600	rgGlnLysProSerAspArgAlaAspSerArgArgSerTyrHisGlnGlnLysProPheG	620
Db	2336	GGCAGAAAGCCAAAGTACAGAGCTGACTCGCGGGAGCTGGCATGAAGAGGCCCTTTG	2395
Qy	620	LysGlnPheLysArgArgSerCysGlnMetGluPheGlyGlnSerIleMetSerGluA	640
Db	2396	AAAAGAGTTTAAACGAGAAAGCTGCCMAATGAAATTTGAGAGAGACATCATGTACAGA	2455
Qy	640	snArgSerArgGlnGlnLeuGlyLysValGlySerGlnSerSerPheSerGlySerMetG	660
Db	2456	ACAGGTACCGGAAAGACTGGGAAAGTGGGCACTGACGTCTTACCTTTTGGGCGAGATGG	2515
Qy	660	lulleIleGluValSer	665
Db	2516	AAATCATTTAGAGTCTCC	2532
RESULT 13			
ID	ABN83966	standard; DNA; 4790 BP.	
AC	ABN83966;		
XX	06-SEP-2002	(first entry)	
XX	Human gene sequence #13.		
DE	Human; brain; tonsil; hippocampus; foetal brain; diagnosis; gene; de.		
XX			
KM			


```
QY 480 InsSerLysArgLeuHisSerValArgThrSerSerSerGlyThrAlaGlnArgSerLeuL 500
XX |||||
XX 1622 AAGAGCAAGCATTTGCTATTCGGTCCAGAACGACGAGCTGGCACCCGCCAGAGGTCCTCTT 1681
Db |||||
QY 500 euSerProLeuHisArgSerGlySerValGluAspAsnThrHisThrSerPheLeuPheG 520
XX |||||
XX 1682 TATCTCCTACCTGCATCCAGAGTGGAGCGGTGAGAGCAATTAACACACACCTTCCTTTTGG 1741
Db |||||
QY 520 lYLeuSerThrSerGlnGlnHisLeuThrLysSerAlaGlyLeuGlyLeuLysGlyTTPH 540
XX |||||
XX 1742 GCCTTCCACCAAGCCAGCAGACCTTCACAGAGTCTGCGCTGGCCTTTAAGGCGTGGC 1801
Db |||||
QY 540 iSerAspIleLeuAlaProGlnThrSerThrProSerLeuThrSerSerTTPYrPhea 560
XX |||||
XX 1802 ACTCGGATATCTTGGCCCCCAGACCTTAACCCCTTCCTCCAGACACACCTGGTATTTTG 1861
Db |||||
QY 560 lATHrGlySerSerHisPheTYrSerAlaSerAlaIleTYrGlyGlySerAlaSerTYrs 580
XX |||||
XX 1862 CCACAGATCCTCACCTTACTCTGCTTCAACCACTTAACGAGGAGGCACTGCCAGTTACT 1921
Db |||||
QY 580 eAlaTYrSerCysSerGlnLeuProThrCysGlyAspGlnValTYrSerValArgArg 600
XX |||||
XX 1922 CTGGCTACAGCTGCACGACGCTGCCACTTGGCGAGACCAAGTCTATTCGTGGCGCAGGC 1981
Db |||||
QY 600 rGlnLysProSerAspArgAlaAspSerArgArgSerTTPHISGlnGlySerProPheG 620
XX |||||
XX 1982 GGCAGAAAGCCAAAGTGCACAGAGCTGACTCGCGGAGGCTGGCATGAAGAGAGACCCCTTTG 2041
Db |||||
QY 620 lLysGlnPheLysArgArgSerCysGlnMetGlnPheGlyGlnSerIleMetSerGln 640
XX |||||
XX 2042 AAAAGAGATTTAACCCAGACGCTGCCAATGGAAATTGGAGAGACATCATCTCAGAGA 2101
Db |||||
QY 640 snArgSerArgGlnGlnLeuGlyLysValGlySerGlnSerSerPheSerGlySerMetG 660
XX |||||
XX 2102 ACAGGTCACGAGGAGAGCTGGGGAAAGTGGGAGCTGAGCTTTTGGGGCAGCATGG 2161
Db |||||
QY 660 lullelleglValSer 665
XX |||||
XX 2162 AATCATTTGAGGCTCC 2178
Db |||||

RESULT 14
ID ABV20833
ABV20833 standard; cDNA; 5145 BP.
XX
AC ABV20833;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 20824.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-018319P.
XX
PR 16-MAR-2000; 2000US-018962P.
XX
PR 25-MAY-2000; 2000US-0207454P.
XX
PR 09-JUN-2000; 2000US-0211314P.
XX
PR 18-JUL-2000; 2000US-021907P.
XX
PR 13-DEC-2000; 2000US-0255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JR;
XX
```

```
DR MPI; 2001-662795/76.
XX
XX PT Novel isolated nucleic acid molecule associated with cancerous state of
XX prostate cells and correlating with presence of prostate cancer; useful
XX for detecting presence of prostate cancer, stage of prostate cancer.
PS Claim 1; Page 3419; 11750pp; English.
XX
XX CC The invention relates to an isolated nucleic acid molecule (I) comprising
XX a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX specification or its complement. (I) is useful for: (a) assessing whether
XX a patient is afflicted with prostate cancer; (b) monitoring the
XX progression of prostate cancer in a patient; (c) assessing the efficacy
XX of a test compound to inhibit prostate cancer in a patient; (d) assessing
XX the efficacy of a therapy for inhibiting prostate cancer in a patient;
XX (e) selecting a composition for inhibiting prostate cancer in a patient;
XX (f) assessing the prostate cell carcinogenic potential of a compound; (g)
XX determining whether prostate cancer has metastasized in a patient; (h)
XX assessing the aggressiveness or indolence of prostate cancer in a patient
XX ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
SQ Sequence 5145 BP; 1408 A; 1135 C; 1253 G; 1346 T; 0 U; 3 Other;
XX
XX Alignment Scores:
XX Pred. No.: 0 Length: 5145
XX Score: 472.00 Matches: 663
XX Percent Similarity: 99.40% Conservative: 0
XX Best Local Similarity: 99.40% Mismatches: 2
XX Query Match: 70.98% Indels: 4
XX DB: 5 Gaps: 0
XX
US-10-029-345A-109 (1-665) x ABV20833 (1-5145)
QY 1 MetAlaHisGlnMetIleGlyThrGlnIleValThrGluArgLeuValAlaLeuLeuGln 20
XX |||||
XX 589 ATGCCCATGAGATGATTGGAATTCGAATTTGACTGAGAGGTTGGTCTCTGTGGAA 648
Db |||||
QY 21 SerGlyThrGlnLysValLeuLeuIleAspSerArgProPheValGlnTYrAsnThrSer 40
XX |||||
XX 649 AGTGAAGCGAAMAAAGCTGCTGAATTGATGACCGGCAATTTGTGAAATACATACATCC 708
Db |||||
QY 41 HisIleLeuGlnAlaIleAsnIleAsnCysSerLysLeuMetLysArgArgLeuGln 60
XX |||||
XX 709 CACATTTTGAAGCATTATATATCACTGCTCCAAAGCTTATGAACGAAGGTGCAACAG 768
Db |||||
QY 61 AspLysValLeuIleThrGlnLeuIleGlnHisSerAlaLysHisLysValAspIleAsp 80
XX |||||
XX 769 GACAAAGTTATTAACAGAGCTCATCCAGATTCCAGCGAAACATTAAGGTTGACATTGAT 828
Db |||||
QY 81 CysSerGlnLysValValValTYrAspGlnSerSerGlnAspValAlaSerLeuSerSer 100
XX |||||
XX 829 TGCAGTCCAGAAAGTTGATGATTACATCAAGCTCCAAAGATGTGCTCTCTCTCTCA 888
Db |||||
QY 101 AspCysPheLeuThrValLeuLeuGlyLysLeuGlnLysSerPheAsnSerValHisLeu 120
XX |||||
XX 889 GACTGTTTTCACCTGATCTTCTGGGTAACCGAAGAGAGCTTCACTCTGTTCACTCG 948
Db |||||
QY 121 LeuAlaGlyGlyPheAlaGlnPheSerArgCysPheProGlyLeuCysGlnGlyLysSer 140
XX |||||
XX 949 CTTCAGAGTGGGTTGCTGAGATTCTCTGTTGTTCCCTGGGCTCTGTGAAGGAAATCC 1008
Db |||||
QY 141 ThrLeuValProThrCysIleSerGlnProCysLeuProValAlaAsnIleGlyProThr 160
XX |||||
XX 1009 ACTTAGATCCCTACCTGCAATTTCTAGGCTTGCTTAACCTGTGGCCAAACCTGGCCAAAC 1068
Db |||||
QY 161 ArgIleLeuProAsnLeuTYrLeuGlyCysGlnArgAspValLeuAsnLysGlnLeuIle 180
XX |||||
XX 1069 CGAATTTTCCCAATCTTATCTTGGCTGCCAGAGATGCTCTCAACAGAGAGCTGAT- 1127
Db |||||
QY 181 GlnGlnAsnGlyIleGlyTYrValLeuAsnAlaSerTYr-THrCysProLysProAsp 200
XX |||||
XX 1128 GCAGCAGATGGAGATTGTTATGTGTAAATGCGACAA-TACCTGTCCAAAGCTGACT 1186
Db |||||
```

QY	200	hellEPRGlnSerHisPheLeuAGValProValAsnAspSerPheCysGlnYsrlleL	220
Db	1187	TTATCCCGAGCTCATATTTCCGTGGTGGCCCTGTAATGACGCTTTGTGAGAAATTT	1248
QY	220	eUPrTriPLeuAspLysSerValAspPheileGluLysAlaLysAlaSerAngLYsV	240
Db	1247	TCGCCGTGTGACAAATCAGTAAATTTTCATTGAGAAAGCAAAAGCTCCATGGATGTG	1306
QY	240	alLeuValHiCysLeuAlaGlylleSerArgSerAlaThrileAlleAlaYrilem	260
Db	1307	TTCTAGTGCACGTTTAACTGGAGATCTCCCGCTCCGACCACTCGCTATCGCTACATCA	1368
QY	260	eLYVAArgMetAspMetSerLeuAspGluAlaTyrArgPheAllyleGluLysAspProT	280
Db	1367	TGAAGAAGATGACATGTCTTTAGATGAAGCTTACAGATTTGTGAAAGAAAAAACCTTA	1428
QY	280	hrliSeSProAsnPheAsnPheLeuGlyGlnLeuLeuAspTyrGluLysLysileYSA	300
Db	1427	CTATATCTCCAAACTTCAATTTTCTGGGCCAACTCTTGACCTATAGAAAGAAATTAAAG	1488
QY	300	enGlnThrGlyAlaSerGlyProLysSerLysLeuLysLeuLysHisLeuGlnLysProA	320
Db	1487	ACCAAGACTGGAGCATCAGGGCCAAAGAGAAACTCAAGCTGCTGCACCTGGAGAAAGCCA	1548
QY	320	enGlnProValProAlaValSerGlnGlyGlnLysSerGluThrProLeuSerProp	340
Db	1547	ATGAACCTGTCCCTCTCTGCTCAGAGGGGGGACAAAGAAAGCAGACGCCCTCACTCAC	1608
QY	340	roCyAlaAspSerAlaThrIserGluAlaGlyGlnArgProAlaHisProAlaSerV	360
Db	1607	CTGTGGCGACCTCTCTACTCAGAGGACAGACAAAGGCCCTGTGATCCCGCACAG	1668
QY	360	aLProSerValProSerValGlnProSerLeuLeuGluAspSerProLeuValGlnAlaL	380
Db	1667	TGCCACGGGTGCCACGCGTGCAGCCGTGTTAGAGACAGCCCGCTGGTACAGGGCG	1728
QY	380	eUSerGlyLeuLysLeuSerAlaAspArgLeuLysAspSerAsnLysLeuLysArgSerP	400
Db	1727	TCAGTGGGTGCACCTGTCTCCGACAGCGCTGGAGAGACAGCATATTAAGCTCAAGCTTCT	1788
QY	400	heSerLeuAspLileYsSerValSerTyrSerAlaSerMetAlaSerLeuHisGlyL	420
Db	1787	TCTCTCTGGATATCAAAATCAGTTTCAATTTACGCCAGCATGGCACATCTTCACTAGCT	1848
QY	420	heSerSerSerGluAspAlaLeuGlnTyrTyrLysProSerThrThrLeuAspGlyThrA	440
Db	1847	TCTCTCATACAAATATGCTTTGGAAATCTACAAACCTTCCACTCTGTGATGGAGCA	1906
QY	440	eNLYLeuLysCysGlnPheSerProValGlnLysLeuSerGlnGlnThrProGlnThrSerP	460
Db	1907	ACAAAGCTATGCGCAGTTCTCCCTGTTACAGAACTATCGAGAGCAGCTCCCGAAACCACTC	1968
QY	460	roAspLysGlnGlnLysSerileProLysLysLeuGlnThrAlaArgProSerAspSerG	480
Db	1967	CTGAATAAAGAGAAACCCAGCACTCCCAAGAACTCAAGCCGCCAGGCTTTCAGCACACC	2028
QY	480	InSerLysArgLeuHisSerValArgThrSerSerSerGlyThrAlaGlnArgSerLeuL	500
Db	2027	AGAGCAAGCATTTGCATTCGGTCAACAAACAGACAGATGGCACGCCCCAGAGGTCCCTTT	2088
QY	500	eUSerProLeuHisArgSerGlySerValGluAspAsnTyrHisThrSerPheLeuPheG	520
Db	2087	TATCTCCACTGCATCGAAGTGGAGACGTGAGAGCAATTACCACACCAAGCTTCCCTTTG	2148
QY	520	lyLeuSerThrSerGlnGlnHisGluThrTrpLysSerAlaGlyLeuGlyLeuLysGlyTTPH	540
Db	2147	GCTTTTCCACCAAGCCAGCAGCACTTCACGAAATGTGCTGGCCCTTAAAGGCTGGC	2208
QY	540	lAsSerAspLileuLysProGlnThrSerThrProSerLeuThrSerSerTTPYrPheA	560
Db	2207	ACTCGGATATCTTGGCCCCCAGACCTCTAACCTTCCCTTACACGACAGCTGATTTTGG	2268
QY	560	lathGlnSerSerHisPheTyrSerAlaSerAlaileTyrGlyGlySerAlaSerTyrS	580

Db	2267	CCAAGAGTCTTCACTTCTACTCTGCTCAGGCATCTACGAGGCGAGTCCAGTTACT	2326
Qy	580	era1aTySerCySserGlnLeuProThrCysglYaaSpGlnValTySerValaArga	600
Db	2237	CTGCTACAGCTGACGACGCTGCCACTTGCGGAGACCAAGTCTATTCTGTGGCAGGC	2386
Qy	600	rgGlnTyPProSerAspArgAlAaSPeSerArgaTgSerTPPhISglGlnSerProheg	620
Db	2387	GGCAGAGCCCAAGTGAAGAGAGCTGACTCGCGCGGAGCTGGCATGMAAGAGCCCCCTTG	2446
Qy	620	IuLySGlnPheLyArgArgSerCySglInMeGluPheglYgluSerTlImerSerGluA	640
Db	2447	AAAGCGAGTTTAAACGACGAAGCTGCGCAATGGAAATTTGGAGACGACATCATGTCAAGA	2506
Qy	640	snArgSerArgGluGlnLeuGlnTyLySerValglYserGlnSerSerPheSerGlySerMeG	660
Db	2507	ACAGTGACCGGGAAGAGCTGGCGAAATGGGCGCATCATGCTTTTCCGGCAGCATGG	2566
Qy	660	IuIleIleGluValSer	665
Db	2567	AAATCATTTGAGTCTCC	2583
RESULT 15			
ABV21080			
ID	ABV21080	standard; cDNA, 5145 BP.	
XX	AC	ABV21080;	
XX	DT	13-SEP-2002 (first entry)	
XX	DE	Human prostate expression marker cDNA 21071.	
XX	XX	Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;	
XX	KW	pharmacogenomic marker; gene; ss.	
XX	OS	Homo sapiens.	
XX	PN	WO200160860-A2.	
XX	PD	23-AUG-2001.	
XX	PF	20-FEB-2001; 2001WO-US0005171.	
XX	PR	17-FEB-2000; 2000US-0183319P.	
XX	PR	16-MAR-2000; 2000US-0189862P.	
XX	PR	25-MAY-2000; 2000US-0207454P.	
XX	PR	09-JUN-2000; 2000US-0211314P.	
XX	PR	18-JUL-2000; 2000US-0219007P.	
XX	PR	13-DEC-2000; 2000US-0255281P.	
XX	PA	(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.	
XX	PI	Schlegel R, Endege WO, Monahan JE;	
XX	DR	WPI; 2001-662795/76.	
XX	PT	Novel isolated nucleic acid molecule associated with cancerous state of	
XX	PT	prostate cells and correlating with presence of prostate cancer, useful	
XX	PT	for detecting presence of prostate cancer, stage of prostate cancer.	
XX	PS	Claim 1; Page 3481; 11750pp; English.	
CC	The invention relates to an isolated nucleic acid molecule (I) comprising		
CC	a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the		
CC	specification or its complement. (I) is useful for: (a) assessing whether		
CC	a patient is afflicted with prostate cancer; (b) monitoring the		
CC	progression of prostate cancer in a patient; (c) assessing the efficacy		
CC	of a test compound to inhibit prostate cancer in a patient; (d) assessing		
CC	the efficacy of a therapy for inhibiting prostate cancer in a patient;		
CC	(e) selecting a composition for inhibiting prostate cancer in a patient;		
CC	(f) assessing the prostate cell carcinogenic potential of a compound; (g)		
CC	determining whether prostate cancer has metastasized in a patient; (h)		

CC assessing the aggressiveness or indolence of prostate cancer in a patient
 CC ; (1) is also useful as a pharmacodynamic or pharmacogenomic marker
 XX
 SQ Sequence 5145 BP; 1408 A; 1135 C; 1253 G; 1346 T; 0 U; 3 Other;

Alignment Scores:

Pred. No.:	0	Length:	5145
Score:	472.00	Matches:	663
Percent Similarity:	99.40%	Conservative:	0
Best Local Similarity:	99.40%	Mismatches:	2
Query Match:	70.98%	Indels:	4
DB:	5	Gaps:	0

US-10-029-345A-109 (1-665) x ABV21080 (1-5145)

```

QY 1 MetAlaHisGluMetIleGlyThrGlnIleValThrGlnArgLeuValAlaLeuLeuGlu 20
DB 589 ATGGCCCATGAGATGATGATGAACTCAAAATTTGATCGAGAGGTGGTGGCTGTGCGTGA 648
QY 21 SerGlyThrGlnuysValLeuLeuIleAspSerArgProPheValGlnuysPheThrSer 40
DB 649 AGTGGAAACGGAAGAAAGTGGCTGCTAATTTGATAGCCGCGCATTTGTGAAATACATACATCC 708
QY 41 HisIleLeuGlnuValIleAsnIleAsnCySerLysLeuMetLysArgArgLeuGlnGln 60
DB 709 CACATTTTGGAGCCATTAATCAACTGCTCCAGCTTATGAAAGCGAAGTTGCAACAG 768
QY 61 AspLysValLeuIleThrGlnuLeuIleGlnHisSerAlaLysHisLysValAspIleAsp 80
DB 769 GACAAAGTTTAATTAACAGACTCATCCAGCATTCAGCAAAACATTAAGGTGACATTTGAT 828
QY 81 CySerGlnuysValValValTyraSpGlnSerSerGlnAspValAlaSerLeuSerSer 100
DB 829 TGGAGTCAGAGAGGTGTAGTTTACATCAAGCTCCAAAGTTCCTCTCTCTCTCA 888
QY 101 AspCyPheLeuThrValLeuLeuGlnLysLeuGlnLysSerPheAsnSerValHisLeu 120
DB 889 GACTGTTTCTCACTGACTCTGGGTTAACTGAGAAAGAGCTTCAACTCTGTCACTGG 948
QY 121 LeuAlaGlyGlyPheAlaGlnPheSerArgCyPheProGlyLeuCyGlnuGlyLysSer 140
DB 949 CTTCGAGGTGGGTTTGGTCTGAGTTCTCTGTTGTTCCCTGGGCTCTGTAAAGAAATCC 1008
QY 141 ThrLeuValProThrCySerIleSerGlnProCySerProValAlaAsnIleGlyProThr 160
DB 1009 ACTCTAGTCCCTACCTGATTTCTGAGCTTCTGCTTACCTGTTGCCAACATTGGGCAACC 1068
QY 161 ArgIleLeuProAsnLeuTyrlLeuGlyCyGlnuArgAspValLeuAsnLysGlnuLeuIle 180
DB 1069 CGAATTTCTTCCAAATCTTATCTTGGCTGCCAGCAGATGCTCAACAAAGAGAGCTGAT- 1127
QY 181 -GlnGlnAsnGlyIleGlyTyrlValLeuAsnAlaSerTyrl-ThrCySPolySProAsp 200
DB 1128 GAGCAGAAATGGGATTTGGTATAGTGTAAATGCCAGCAA-TACCTGTCCAAAGCCGTGACT 1186
QY 200 IleIleProGlnuSerHisPheLeuArgValProValAsnAspSerPheCyGlnuTyrlLeu 220
DB 1187 TTATCCCGGAGTCTCATTTCTCGCGTGTGCTGTGAATGACAGCTTTTGTGAGAAATTT 1246
QY 220 euProThrLeuAspLysSerValAspPheIleGlnuLysAlaLysAlaSerAsnGlyCySV 240
DB 1247 TGGCGTGTGTGCAAAATCAAGATTTTCATTGAAAGAAAGCAAAAGCCCTCAATGATGATGTG 1306
QY 240 AlLeuValHisCySeruAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrlIleM 260
DB 1307 TTCTAGTGCACGTGTTTACGTGGGATCTCCCGCTCCGCCAACATTCCTATTCGCTTACATCA 1366
QY 260 eLysArgMetAspLysSerLeuAspGlnuAlaTyrlArgPheValLysGlnuLysArgProT 280
DB 1367 TGAAGAGATGAGCATGTCTTATGATGAAGCTTACAGATTTGTGAAAGAAAGAAAGACTTA 1426
QY 280 hrIleSerProAsnPheAsnPheLeuGlyGlnLeuLeuAspTyrlGlnuLysLysIleLysA 300

```

```

DB 1427 CTATATTCCAAACTTCAATTTTCGGGCCCACTCTGGACTATGAGAAAGATTAAAGA 1486
QY 300 enGlnThrGlyAlaSerGlyProLysSerLysLeuLysLeuLeuHisLeuGlnuLysProA 320
DB 1487 ACCAGACTGGAGCATCAGGCGCCAAAGACAAACTCAAGCTGTGACCTGGAGAGGCCAA 1546
QY 320 enGlnProValProIleValSerGlnuGlyGlnLysSerGlnuThrProLeuSerProP 340
DB 1547 ATGAACCTGTCCCTGTCTCTCAGAGGAGTGAACGAAAGAGAGAGCGCCCTCAGTCCAC 1606
QY 340 roCyAlaAspSerAlaThrSerGlnuAlaIleGlyGlnArgProValHisAspAlaSerV 360
DB 1607 CCTGTGCCGACTGTCTACTCTCAGAGGAGCAGACGAAAGGCCCTGTGATCCCGCAGACG 1666
QY 360 AlProSerValProSerValGlnProSerLeuLeuGlnuAspSerProLeuValGlnuAla 380
DB 1667 TGGCCAGCGTGGCCAGCTGACGCGCTGTGATTAGAGAGACAGCCCGCTGTACAGGCGC 1726
QY 380 euSerGlyLeuHisLeuSerAlaAspArgLeuGlnuAspSerAsnLysLeuLysArgSerP 400
DB 1727 TCAGTGGGCTGACCTGTCCGACAGAGGCTGAGAAAGACGAAATTAAGCTCAAGGCTTCT 1786
QY 400 heSerLeuAspIleLysSerValSerTyrlSerAlaSerMetAlaAlaSerLeuHisGlyP 420
DB 1787 TCTCTCTGATATCAAAATCAGTTTCAATTTACAGCCAGATGGAGCATCTTACATGGCT 1846
QY 420 heSerSerSerGlnuAspAlaLeuGlnuTyrlTyrlSProSerThrThrLeuAspGlyThra 440
DB 1847 TCTCTCATCAGAAAGATGCTTTGGAAATTAATAAACTTCCATCTGTGATGGAGCA 1906
QY 440 snLysLeuCyGlnuPheSerProValGlnuLeuSerGlnuGlnuThrProGlnuThrSerP 460
DB 1907 ACAAGCTATGCACTTCTCCCTTTTCAAGAACTATGAGAGACACTCCGAAACCAAGTC 1966
QY 460 roAspLysGlnuGlnuAlaSerIleProLysLysLeuGlnuThrAlaArgProSerAspSerG 480
DB 1967 CTGATTAAGAGAGAGAGCCAGCATCCCAAGAAAGTCGAGACCGGACCTTCAAGAGCC 2026
QY 480 InsSerLysArgLeuHisSerValArgThrSerSerSerGlyThrAlaGlnuArgSerLeu 500
DB 2027 AGAGCAAGCAATTCATTCGCTGTCAGAACAGACAGATGGCAGCCGCAAGGCTCCCTTT 2086
QY 500 euSerProLeuHisArgSerGlySerValGlnuAspAsnTyrlSerPheLeuPheG 520
DB 2087 TATCTCAGCTGCATCGAAGTGGAGGCTGAGAGCAATTCACACAGCTTCTTTTCG 2146
QY 520 LysLeuSerThrSerGlnuGlnuHisLeuThrLysSerAlaGlyLeuGlyLysGlyTyrlP 540
DB 2147 GCCTTTCCACAGCCAGCAGACACCTCAGAAAGTCTGTGGCTTGAAGGCTGGC 2206
QY 540 isSerAspIleLeuAlaProGlnuThrSerThrProSerLeuThrSerSerTyrPheA 560
DB 2207 ACTGGATATCTTGGCCCCCAGACCTTACCCCTTCCGACACAGCAGCTGTGATTTTG 2266
QY 560 IaThrGlnuSerSerHisPheTyrlSerAlaSerAlaIleTyrlGlyLysSerAlaSerTyrlS 580
DB 2267 CCAAGAGTCTCTCACTTTACTCTGCTCTCAACCATCTACGAGGAGAGTCCAGTTACT 2326
QY 580 eAlaTyrlSerCySerGlnuLeuProThrCyGlnuAspGlnuValTyrlSerValArgArgA 600
DB 2327 CTGCTTACAGCTGACCCAGCTGCCACTTGGCAGAGCAAGTCTATTCTGTGCGCAGGC 2386
QY 600 rGlnuLysProSerAspArgAlaAspSerArgArgSerTyrPheIleGlnuLysSerProPheG 620
DB 2387 GGCAGAAAGCCAAAGGAGAGAGTGAATCGGGGAGCTGGCAATGAAGAGAGCCCTTTTG 2446
QY 620 LuLysGlnuPheLysArgArgSerCyGlnuMetGlnuPheGlyGlnuSerIleMetSerGlnuA 640
DB 2447 AAAAGAGTTTAAACGAGAAAGCTCCCAATGAAATTTGAGAGAGACATATTCAGAGA 2506
QY 640 snArgSerArgGlnuGlnuLeuGlyLysValGlySerGlnuSerSerPheSerGlySerMetG 660
DB 2507 ACAGGTCAACGAGAAAGCTGGGAAAGTGGGAGTCAAGTCACTTTTCCGGCAGCATGG 2566

```

QY 660 lulellegluValSer 665
DB 2567 AAATCATTGAGGTCTCC 2583

RESULT 16

ID ABV26680 standard; cDNA; 5145 BP.
XX ABV26680;

AC ABV26680;

DT 16-SEP-2002 (first entry)

DE Human prostate expression marker cDNA 26671.

KM Human, prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX pharmacogenomic marker; gene; ss.

OS Homo sapiens.

XX MO200160860-A2.

XX 23-AUG-2001.

PF 20-FEB-2001; 2001MO-US005171.

XX 17-FEB-2000; 2000US-0183319P.

PR 16-MAR-2000; 2000US-0189862P.

PR 25-MAY-2000; 2000US-0207454P.

PR 09-JUN-2000; 2000US-0211314P.

PR 18-JUL-2000; 2000US-0219007P.

PR 13-DEC-2000; 2000US-0255281P.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

PI Schlegel R, Endege WO, Monahan JE;

XX WPI; 2001-662795/76.

PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.

XX Claim 1; Page 5388-5389; 11750pp; English.

CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV0010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC assessing whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or incidence of prostate cancer in a patient
CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker

XX Sequence 5145 BP; 1408 A; 1135 C; 1253 G; 1346 T; 0 U; 3 Other;

XX Alignment Scores:

Pred. No.: 0 Length: 5145

Score: 472.00 Matches: 663

Percent Similarity: 99.40% Conservative: 0

Best Local Similarity: 99.40% Mismatches: 2

Query Match: 70.98% Indels: 4

DB: 5 Gaps: 0

US-10-029-345A-109 (1-665) x ABV26680 (1-5145)

QY 1 MetAlHisGluMetIleGlyThrGlnIleValThrGlnArgLeuValAlaLeuGlu 20
DB 589 ATGGCCATGATGATGGAATCTCAATTTGTTACTGAGAGTTGGTGGCTCTGCTGAA 648

QY 21 SerGlyThrGluValLeuLeuIleAspSerArgProPheValGluTyrAsnThrSer 40
DB 649 AGTGAACGGAAAAATGCTGCTAATTGATAGCGGCCATTGTTGGAAATCAATATCATCTCC 708

QY 41 HisIleLeuGluAlaIleAsnIleAsnGlySerLeuMetLeuValArgArgLeuGln 60
DB 709 CACATTTTGGAGCCATTATATCACTGCTCCAACTTATGAAAGGAAAGTTGGCAACAG 768

QY 61 AsplValLeuIleThrGluLeuIleGlnHisSerAlaValHisValAspIleAsp 80
DB 769 GACAAAGTGTATTACAGAGCTCATCCAGCATTCAGCAAAACATAGAGTTGACATTGAT 828

QY 81 CysSerGluValValValTyrAspGlnSerSerGlnAspValAlaSerLeuSerSer 100
DB 829 TGCAGTCAGAGGTGTAGTTTACATCAAGCTCCCAAGATGTCCCTCTCTCTCTCA 868

QY 101 AspCysPheLeuThrValLeuLeuGlyLysLeuGluLysSerPheAsnSerValHisLeu 120
DB 889 GACTGTTTCTACCTGCTTCTGAGTAAACTGGAGAGAGCTTCACTCTGTCACCTG 948

QY 121 LeuAlaGlyGlyPheAlaGluPheSerArgCysPheProGlyLeuCysGluGlySer 140
DB 949 CTTCAGGTGGGTTCGATGCTCTCGTGTCTTCCCTGCTCTGTAAGGAAATCC 1008

QY 141 ThrLeuValProThrCysIleSerGlnProCysLeuProValAlaAsnIleGlyProThr 160
DB 1009 ACTTATGCTCCCTACCGCATTTCTCAGCCTTGCTTACCTGTCACCAATTTGGCCAAAC 1068

QY 161 ArgIleLeuProAsnLeuTyrLeuGlyCysGlnArgAspValLeuAsnIleGlyLeu 180
DB 1069 CGAATTCCTCCAAATTTTATCTTGCTGCCAGCAGAGATGCTTCAACAGAGCTGAT 1127

QY 181 -GlnGlnAsnGlyIleGlyTyrValLeuAsnAlaSerTyr-ThrCysProLysProAsp 200
DB 1128 GCAGCAGATGGGATGTTATGTATGTTAATGCCACAA-TACCTGTCCAAAGCTGACT 1186

QY 200 helLeuProLysSerHisPheLeuArgValProValAsnAspSerPheCysGluValIle 220
DB 1187 TTATCCCGCAGCTCTCAATTCCTGCGTGTGCTGTGATACAGCTTTTGGAAAAATTT 1246

QY 220 euProTyrLeuAspLysSerValAspPheIleGluValAlaValAsnSerGlyCys 240
DB 1247 TGCCTGCTTGGACAAATCATGATGATTTTCAATTGAGAAACCAAGCTTCAATGATGTG 1306

QY 240 AlLeuValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIle 260
DB 1307 TTCTAGTCACATGTTTAGCTGGGATCTCCGCTCCGCAACATCGCTATCGCTACATCA 1366

QY 260 eLysArgMetAspMetSerLeuAspGluAlaTyrArgPheValLysGluValArgPro 280
DB 1367 TGAAGAGGATGACATGCTTTTATGAGAGCTTACAGATTTTGGAAAAAGAACCTTA 1426

QY 280 hrIleSerProAsnPheAsnPheLeuGlyGluLeuAspTyrGluValLysIleVal 300
DB 1427 CTATATCTCAAACTTCAATTTCTGGGCCAATCTTGAACATGAGAAAGATTAAGA 1486

QY 300 enGlnThrGlyAlaSerGlyProLysSerLysLeuLysLeuHisLeuGluValPro 320
DB 1487 ACCAGACTGGAGCATGAGGCCCAAGAGCAAACTCAAGCTGTGCGACCTGGAGAACCA 1546

QY 320 enGluProValProAlaValSerGluGlyGlnLysSerGluThrProLeuSerPro 340
DB 1547 ATGAACCTGTCTCTGTCTGTCAGAGGGTGGACAGAAAAAGCAGAGCCCTCAGTCCAC 1606

QY 340 roCysAlaAspSerAlaThrSerGluAlaAlaGlyGlnArgProValHisProAlaSer 360
DB 1607 CCTGTGCGACTGTGCTACCTAGAGGCGAGCGAGCAAAAGGCCCGGTGATCCGCCAGG 1666

QY 360 alProSerValProSerValGlnProSerLeuLeuGluAspSerProLeuValGlnAla 380
DB 1667 TGCCACGCTGCGCCAGCGTGCAGCGCTGCTTGAAGAGACAGCCCGCTGTACAGGCC 1726

```

QY 380 euSerGlyLeuHISeuSerAlaAspArgLeuGluAspSerAnlySerAnlySerArgSerP 400
DB 1727 TCAGTGGGGTGCACCTGTCGCGAGACAGGCTGGAGACAGCAATAGCTCAAGCGCTTCT 1786
QY 400 heSerLeuAspIlelySerValSerTyrSerAlaSerMetAlaAspLeuHISGLYP 420
DB 1787 TCTCTGTGATATCAAAATCAAGTTTCAATATTGAGCCGAGCTGGAGAGATCTTCAACAGCT 1846
QY 420 heSerSerSerGluAspAlaLeuGluTyrTyrLeuProSerThrThrLeuAspGlyThrA 440
DB 1847 TTTCTCTATCAAGAGATGCTTTGGATCTTCAAACTTCTCACTCTGATGGAGCA 1906
QY 440 snlySerCysGlnPheSerProValGlnGluLeuSerGluGlnThrProGluThrSerP 460
DB 1907 ACAAGTATGACAGTTCTCCCTGTTCAAGAACTATCGAGAGACTCCGAAACAGAGTC 1966
QY 460 roAspLySGluGluAlaSerTyrLeuGlyLeuGlnThrAlaArgProSerArgSerG 480
DB 1967 CTGATATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2026
QY 480 InSerLyAspLeuHISeuSerValArgThrSerSerSerGlyThrAlaGlnArgSerLeu 500
DB 2027 AAGAGAGAGATGATGCTGTCGAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2086
QY 500 euSerProLeuHISArgSerGlySerValGluAspAnlyThrIsthrSerPheLeuPheG 520
DB 2087 TATCTTCACTGCATGCAAGTGGAGAGTGGAGAGAGATTAACCAACAGCTTCTTTTGG 2146
QY 520 llyLeuSerThrSerGlnGlnHISleuThrLysSerAlaGlyLeuGlyLeuGlyTyrP 540
DB 2147 GCTTTTCCACAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2206
QY 540 iSerAspIleLeuAlaProGlnThrSerThrProSerLeuThrSerSerTyrPheA 560
DB 2207 ATCTGGATATCTTGGGCCCCCGAGAGCTTCACTTCTTCTGAGCCAGAGCTGTATTTTG 2266
QY 560 laThrGlnSerSerHISpheTyrSerAlaSerAlaIleTyrGlyGlySerAlaSerTyrS 580
DB 2267 CCACAGAGTCTCACACTTCTACTGCTGCTCAGGAGATTAACGAGAGAGAGAGAGAGAGT 2326
QY 580 eAlaTyrSerCysSerGlnLeuProThrCysGlyAspGlnValTyrSerValArgArgA 600
DB 2327 CTGCTTACAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2386
QY 600 rGlnLysProSerAspArgAlaAspSerArgArgSerTrpHisGlnGlnSerProPheG 620
DB 2387 GGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2446
QY 620 llyLeuGlnPheLysArgArgSerCysGlnMetGluPheGlyGlnSerIleMetSerGluA 640
DB 2447 AAAGAGAGTTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2506
QY 640 snArgSerArgGluGluLeuGlyLysValGlySerGlnSerSerPheSerGlySerMetG 660
DB 2507 AAGAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2566
QY 660 llyIleIleGluValSer 665
DB 2567 AAATCATTAAGAGTCTCC 2583

```

```

OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-018313P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PI Schlegel R, Endege WO, Monahan JE;
PI WPI; 2001-662795/76.
DR
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS Claim 1; Page 3451; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
SQ Sequence 5145 BP, 1408 A; 1135 C; 1253 G; 1346 T; 0 U; 3 Other;
XX
Alignment Scores:
Pred. No.: 0 Length: 5145
Score: 472.00 Matches: 663
Percent Similarity: 99.40% Conservative: 0
Best Local Similarity: 99.40% Mismatches: 2
Query Match: 70.98% Indels: 4
DB: 5 Gaps: 0
US-10-029-345A-109 (1-665) x ABV20978 (1-5145)
QY 1 MetAlaHISGluMetIleGlyThrGlnIleValThrGluArgLeuValAlaLeuLeuGlu 20
DB 589 ATGGCCCAATGAGATGATGGAATCAAAATTTTCTGAGAGAGTGGTGGCTCTGCGTGA 648
QY 21 SerGlyThrGluLysValLeuLeuIleAspSerArgProPheValGluTyrAsnThrSer 40
DB 649 AGTGAACAGGAAAGAGTCTCTAATTCATAGCCGCCCATTTTGGAGATCAATACATCC 708
QY 41 HisIleLeuGluAlaIleAsnIleAsnCysSerIysLeuMetLysArgArgLeuGln 60
DB 709 CACATTTTGAAGAGCATTAATATCAAGTCTCCAAGCTTATGAAGCGAAGGTTCGACAG 768
QY 61 AspLysValLeuIleThrGluLeuIleGlnHisSerAlaLysLysValAspIleAsp 80
DB 769 GACAAAGTTAATTAACAGAGCTCATCCAGATTCAGGAAACATTAAGTTGACATTGAT 828
QY 81 CysSerGlnLysValValValTyrAspGlnSerSerGlnAspValAlaSerLeuSerSer 100
DB 829 TGCAGTCAAGAGGTTGTATGTTTACATCAAGAGCTCCCAAGATGTTGCTCTCTCTTCA 888
QY 101 AspCysPheLeuThrValLeuLeuGlyLysLeuGlnLysSerPheAsnSerValHisLeu 120

```

RESULT 17
 ABV20978
 ID ABV20978 standard; cDNA; 5145 BP.
 XX
 AC ABV20978;
 XX
 DT 13-SEP-2002 (first entry)
 XX
 DE Human prostate expression marker cDNA 20969.
 XX
 KM Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 XX

Db 889 GACTGTTTCTCACTTACTTCTGGGTAAGTGAAGAGCTTCAACTCTGTTCACTCG 948
 Qy 121 LeuAlaGlyGlyPheAlaGluPheSerArgPheProGlyLeuGlyGlyLeuSer 140
 Db 949 CTGGCAGGGTGGGTTTCTGAGTCTCTCGTTGTTTCTCGGCTCTGTAAGAGAAATATC 1008
 Qy 141 ThrLeuValProThrCysAlaIleSerGlnProCysLeuProValAlaAsnIleGlyProThr 160
 Db 1009 ACTCTAGTCCCTACCTGCACTTCTCAGCTTCTGCTTACCTGTTGCAACATGGGCAACC 1068
 Qy 161 ArgIleLeuProAsnLeuTyrlleuGlyCysGlnArgAspValLeuAsnIleGlyLeuIle 180
 Db 1069 CGAATTCCTCCCAATCTTTATCTTGGCTGCGAGCGAGATGCTCAACAAGAGCTGAT- 1127
 Qy 181 -GlnGlnAsnGlyIleGlyTyrlleuAlaAsnIleSerTyrlleuProGlyProAsp 200
 Db 1128 GCAGCAGAAATGGATGGTTATGTGTTAAATGCCAGCAA-TACCTGTCCAAAGCTGACT 1186
 Qy 200 heIleProGluSerHisPheLeuArgValProValAsnAspSerPheCysGluValIleL 220
 Db 1187 TTATCCCCCAAGTCTCATTTCTGCTGCTGCTGTGATGACAGCTTTGTGAGAAAATTT 1246
 Qy 220 euProTrpLeuAspIleSerValAspPheIleGluIleValAlaValAsnGlyCysArg 240
 Db 1247 TGCCTGCTGGTGGCAATCAGTAGATTTCATTGAGAAAGCAAAAGCTCCAAATGATGTG 1306
 Qy 240 alleuValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrlleu 260
 Db 1307 TTCTACTGCACTGTATTAGCTGGGATCTCCGCTCCCAACATCGCTATGCTCATCA 1366
 Qy 260 ellyAsArgMetAspMetSerLeuAspGluAlaTyrlleuPheValIleValGluValArgPro 280
 Db 1367 TGAAGGAGATGACATGCTTTAGATGAGCTTACGATTTGTGAAGAAAAAGACCTTA 1426
 Qy 280 hrIleSerProAsnPheAsnPheLeuGlyGlnLeuLeuAspTyrlleuValIleValAsp 300
 Db 1427 CTRATCTCCAACTTCAATTTTCTGGGCCAATCTCTGCACTATGAGAAAGATTTAA 1486
 Qy 300 enGlnThrGlyAlaSerGlyProIleSerIleLeuIleValLeuGlyIleValPro 320
 Db 1487 ACCAGACTGGAGATCAGGGCCAAAGCAAACTCAAGCTGCTGACCTGAGAAACCAA 1546
 Qy 320 enGluProValProAlaValAsnGlyGlyGlnIleSerGlyIleThrProLeuSerPro 340
 Db 1547 ATGAACCTGCTCTGCTGCTCAGAGGAGGAGCAAAAGCGAGCGCCCTCAAGTCCAC 1606
 Qy 340 roCysAlaAspSerAlaThrSerGlyAlaAlaGlyGlnArgProValHisProAlaSer 360
 Db 1607 CCGTGTCCCACTTCTGCTACTCTCAGAGGCAAGCAAAAGCGCCGCTGATCCGCAAGC 1666
 Qy 360 alProSerValProSerValGlnProSerLeuLeuGluAspSerProLeuValGlnAla 380
 Db 1667 TGCCCAAGCGGCGCCAGCGTGCAGCCGTCGTTAGAGGACAGCCGCTGTACAGCGCG 1726
 Qy 380 euSerGlyLeuHisLeuSerAlaAspArgLeuGluAspSerAsnIleValIleValArgSer 400
 Db 1727 TCAGTGGGCTGCACCTGTCCGAGACAGGCTGGAACAGCAATTAAGCTCAACCTTCT 1786
 Qy 400 heSerLeuAspIleLeuSerValSerTyrlleuAlaSerMetAlaAlaSerLeuHisGly 420
 Db 1787 TCTCTTGATATCAATCAATCAATTCATTCACCGACATGGACATCTTCAATGAGCT 1846
 Qy 420 heSerSerSerGluAspAlaLeuGluTyrlleuValProSerThrThrLeuAspGlyThr 440
 Db 1847 TCTCCATCATCAGAAAGATGCTTGAATATCAAAACCTTCACTGATCGAGTGGACCA 1906
 Qy 440 enIleLeuCysGlnPheSerProValGlnIleLeuSerGlyGlnThrProGluThrSer 460
 Db 1907 ACAAGCTATGCGAGTCTCCCTGTTCAAGAACTATCGAGACAGACTCCCGAAACCAAG 1966
 Qy 460 roAspIleGluGluAlaSerIleProIleValLeuGlnThrAlaArgProSerAspSer 480

Db 1967 CTGATAGAGAGAACCCAGCATCCCAAGAACTGCAGACCCGACGCTTGCAGACGCC 2026
 Qy 480 InsIleValArgLeuHisSerValArgThrSerSerSerGlyThrAlaGlnArgSerLeu 500
 Db 2027 AAGCAGAGGATGTGATTCGTGCAAGAACGAGCAGTGGCAACCGCCAGAGGTCCTTT 2086
 Qy 500 euSerProLeuHisArgSerGlySerValGluAspAsnIleThrIleSerPheLeuPhe 520
 Db 2087 TATCTCCATGCAATGCAAGTGGAGCGTGGAGAGACATTAACACACCACTTCTTTTG 2146
 Qy 520 IlyLeuSerThrSerGlnGlnHisLeuThrIleSerAlaGlyLeuGlyLeuValGlyTrp 540
 Db 2147 GCCTTCCACAGCAGCAGCAGCAGCTCAAGAACTGCTGCTGGGCTTAAAGGCTGGC 2206
 Qy 540 IAspSerAlaLeuAlaProGlnThrSerThrProSerIleuThrSerSerTrpIlePhe 560
 Db 2207 ACTCGATATCTTGGCCCCCGACCTTACCTTCCCTGACAGCAGCAGCTGATTTTG 2266
 Qy 560 IaThrGluSerSerHisPheTyrlleuSerAlaSerAlaIleTyrlleuGlySerAlaSer 580
 Db 2267 CACAGAGTCTCACACTTCTACTGCTGCTCAGCCATCTACGAGGACATGCCAGTTACT 2326
 Qy 580 exAlaTyrlleuSerCysSerGlnLeuProThrCysGlyAspGlnValTyrlleuValArg 600
 Db 2327 CTGCTTACAGCTGACAGCAGCTGCCCACTTGGGAGAACCAAGTCTATCTGTGCGCAG 2386
 Qy 600 rGlnIleProSerSerAspAlaAspSerArgArgSerTrpHisGluGluSerProPheG 620
 Db 2387 GGCAGAGCAGAGTGCAGAGCTGACTCGCGGAGAGCTGGGATGAAAGAGACCCCTTG 2446
 Qy 620 IuIleGlnPheIleValArgArgSerCysGlnMetGluPheGlyGluSerIleMetSerGlu 640
 Db 2447 AAAAGCAGTTTAAAGCAGAACTGTCCAATGAAATTTGAGAGACATCATGTCAGAGA 2506
 Qy 640 snArgSerArgIleGluLeuGlyIleValGlySerGlnSerSerPheSerGlySerMet 660
 Db 2507 ACAGGTCAAGGGAAGAGCTGGGGAAGTGGGAGTCAAGTCAAGTCTTTTGGGAGCATG 2566
 Qy 660 IuIleIleGluValSer 665
 Db 2567 AAATCATGAGGCTTCC 2583
 RESULT 18
 ABV21092
 ID ABV21092 standard; cDNA; 5145 BP.
 AC ABV21092;
 XX
 DT 13-SEP-2002 (first entry)
 XX
 DE Human prostate expression marker cDNA 21083.
 XX
 DE Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200160860-A2.
 XX
 PD 23-AUG-2001.
 XX
 PF 20-FEB-2001; 2001WO-US005171.
 XX
 PR 17-FEB-2000; 2000US-018319P.
 PR 16-MAR-2000; 2000US-0189862P.
 PR 25-MAY-2000; 2000US-0207454P.
 PR 09-JUN-2000; 2000US-0211314P.
 PR 18-JUL-2000; 2000US-0219007P.
 PR 13-DEC-2000; 2000US-0255281P.
 XX
 PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
 XX
 PI Schlegel R, Endege WO, Monahan JB;

XX WPI; 2001-662795/76.

PT Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer.

PS Claim 1; Page 3485; 11750pp; English.

XX The invention relates to an isolated nucleic acid molecule (1) comprising a nucleotide sequence given in Tables 1-9 (ABV0010-ABV6213) of the specification or its complement. (1) is useful for: (a) assessing whether a patient is afflicted with prostate cancer; (b) monitoring the progression of prostate cancer in a patient; (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient; (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient; (e) selecting a composition for inhibiting prostate cancer in a patient; (f) assessing the prostate cell carcinogenic potential of a compound; (g) determining whether prostate cancer has metastasized in a patient; (h) assessing the aggressiveness or indolence of prostate cancer in a patient; (i) is also useful as a pharmacodynamic or pharmacogenomic marker

XX Sequence 5145 BP; 1408 A; 1135 C; 1253 G; 1346 T; 0 U; 3 Other;

Alignment Scores:

Pred. No.:	0	Length:	5145
Score:	472.00	Matches:	663
Percent Similarity:	99.408	Conservative:	0
Best Local Similarity:	99.408	Mismatches:	2
Query Match:	70.988	Indels:	4
DB:	5	Gaps:	0

US-10-029-345A-109 (1-665) x ABV21092 (1-5145)

QY	1	MetAlaHisGluMetIleGlyThrGlnIleValThrGluArgLeuValAlaLeuLeuGlu	20
DB	589	ATGGCCCAAGATGATGGAATCTCAATTTGTTACTGAGAGGTGGTGGCTCTGCTGGAA	648
QY	21	SerGlyThrGluValLeuLeuIleAspSerArgProPheValGluTyrAsnThrSer	40
DB	649	AGTGGACGGAAGAGTGTGCTAAATGATAGCCGGCCATTGTGGAATACAAATCATCC	708
QY	41	HisIleLeuGluAlaIleAsnIleAsnGlySerLeuMetLeuArgArgLeuGlnGln	60
DB	709	CACATTTTGGAGCCATTAAATCACTGCTCCAGCTTATGACGGAAGTTGCAACG	768
QY	61	AspIleValLeuIleThrGluLeuIleGlnHisSerAlaIleHisIleValAspIleAsp	80
DB	769	GACAAAGTGTAAATTAACAGAGCTCATCCAGCATTCAGCGAAACATTAAGTTGACATTGAT	828
QY	81	CysSerGlnIleValValTyrAspGlnSerSerGlnAspValAlaSerLeuSerSer	100
DB	829	TGCAGTCAGAGAGTTGTAATTAAGATCAAAAGCTCCCAAGATGTTGCCCTCTCTTCA	888
QY	101	AspCysPheLeuThrValLeuLeuGluIleGluIleGluIleGluIleGluIleGluIle	120
DB	889	GACTGTTTCTCAGCTGACTCTCTGGGTAACCTGGAGAGAGCTTCACTCTGTTCACTG	948
QY	121	LeuAlaGlyIlePheAlaGluPheSerArgCysPheProGlyLeuCysGluGlyIleSer	140
DB	949	CTTGAGGAGGTGGTGTGAGTTCCTCTGTTGTTCCCTGGGCTCTGTGAAGGAATACC	1008
QY	141	ThrLeuValProThrCysIleSerGlnProCysLeuProValAlaAsnIleGlyProThr	160
DB	1009	ACTTAGTCCCTACCTGCAATTTCTCAGCTTCTTAACCTGTTGCCAACAATGGGCAACC	1068
QY	161	ArgIleLeuProAsnLeuTyrLeuGluCysGlnArgAspValLeuAsnIleGluLeuIle	180
DB	1069	CGAATTCCTCCCAATCTTATCTTGGCTGGCCAGAGAGTCTCTCAACAGAGAGTGAAT	1127
QY	181	GlnGlnAsnGlyIleGlyTyrValLeuAsnAlaSerTyr-ThrCysProIleProAsp	200
DB	1128	GCAGCAGATGAGATGTTATGTATGTAAATGCCAGCAA-TACCTGTCCAAAGCTGACT	1186

QY	200	heIleProGluSerHisPheLeuArgValProValAsnAspSerPheCysGluIleIle	220
DB	1187	TTATCCCGAGCTATTTCTGGCTGTGCTGTAAGAGCTTTGTGTGAAATTT	1246
QY	220	eupProTyrLeuAspIleSerValAspPheIleGluIleGluIleGluIleGluIleGluIle	240
DB	1247	TGCCGTGTTGGACAAATCAGTAAATTTCAATTGAGAAAGCAAAAGCTCCCATATGATGTG	1306
QY	240	AlleuValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaIleAlaIle	260
DB	1307	TTCTAGTCACTGTTTACTGCTGGATCTCCCTCCCTCCGACCATGCTTCCCTACATCA	1366
QY	260	eIleValMetAspMetSerLeuAspGluIleTyrArgPheValIleGluIleValAspProT	280
DB	1367	TGAAGAGATGACATGCTTATGAGAGCTTACAGATTTGTGAAAGAAAGAAAGAAAGCTTA	1426
QY	280	hrlIleSerProAsnPheAsnPheLeuGluGlnIleLeuAspTyrGluIleGluIleGluIle	300
DB	1427	CTATATCTCCAAACTTCAATTTCTGGGCCAACTCTGACTATGAGAAAGAAAGTTAAGA	1486
QY	300	enGlnThrGlyAlaSerGlyProIleSerIleValLeuIleHisLeuGluIleValProA	320
DB	1487	ACCGAGCTGAGCATCAGGGCCAAAGCAAACTCAAGCTGCTGCACTGAGAGAGCCAA	1546
QY	320	enGluProValProAlaValSerGluIleGluIleValSerGluThrProLeuSerProP	340
DB	1547	ATGAACCTCTCCCTGCTGTCTCAAGAGGTGACAGAAAGCAAGAGCCCTCAGTCCAC	1606
QY	340	roCysAlaAspSerAlaThrSerGluAlaIleGluIleGluIleGluIleGluIleGluIle	360
DB	1607	CCTGTGGCAGCTGCTATCTCAGAGGACAGCAAGAGCCCGTGTGATCCCGCCAGCG	1666
QY	360	alProSerValProSerValGlnProSerLeuLeuGluIleValSerProLeuValGlnAla	380
DB	1667	TGCCAGGCTGCCAGCGCTGACGCGCTGTTAGAGAGCAGCCGCTGTTAAGAGCGC	1726
QY	380	eusSerGlyLeuHisLeuSerAlaAspArgLeuGluIleAspSerAsnIleValIleValIle	400
DB	1727	TGAGTGGCTCAGCTGCTGCCAGACAGCTGGAAGACAGCAATTAAGCTCAAGGCTTCT	1786
QY	400	heserLeuAspIleIleSerValSerTyrSerAlaSerMetAlaIleSerLeuHisGlyP	420
DB	1787	TCTCTCGGATATCAATCAATTCATATTCAGCAGATGGACATCTTACATGCT	1846
QY	420	heserSerSerGluAspAlaLeuGluTyrTyrIleProSerThrThrLeuAspGlyThrA	440
DB	1847	TCTCTCATCAGAAAGATGCTTGGAAATTAATAAACTTCCACTTCTGATGGAGCA	1906
QY	440	snIleValCysGlnPheSerProValGlnIleValSerGluGlnThrProGluThrSerP	460
DB	1907	ACAAAGCTATGCCAGTCTTCCCTGTGAGAACTATGAGACAACTCCGAAACCAAGTC	1966
QY	460	roAspIleGluGluAlaSerIleProIleValSerIleGluIleGluIleGluIleGluIle	480
DB	1967	CTGATTAAGAGAGAGCCAGATCCCAAGAAAGCTGACAGCCGACGCTTCAACAGACC	2026
QY	480	IleSerIleValGluHisSerValArgThrSerSerSerGlyThrAlaGlnArgSerLeu	500
DB	2027	AGACCAAGCGATTCATTCGATCGTCAAGAACAGACAGAGTGGACCGCCAGAGGTCTTT	2086
QY	500	eusSerProLeuHisArgSerGlySerValGluIleAspAsnTyrHisThrSerPheLeuPhe	520
DB	2087	TATCTCACTGCAATCAAGAGTGGAGGCGTGAAGCAATTAACACACAGCTTCTTTTCG	2146
QY	520	IleSerThrSerGlnGlnHisLeuThrIleValSerAlaGlyIleGluIleGluIleGluIle	540
DB	2147	GCCTTCCACAGCAGCAGACCTCAAGAAAGTGTGCTGGCTGGGCTTAAGAGGCTGGC	2206
QY	540	IleSerAspIleLeuAlaProGlnThrSerThrProSerLeuThrSerSerTyrTyrPheA	560
DB	2207	ACTGGAATATCTGGCCCGCCAGACCTTACCCCTTCCGACAGCAGCTGATATTTTG	2266

QY 560 lathrGluserSerhlspherYSerAlaserAlaierYrglyGlySerAlaserTyrS 580
DB 2267 CCACAGAGCTCTCACCTTCTACTCTGCTCAGCCATCTACGAGGACAGTCCAGTTACT 2326
QY 580 errAlaYrSerCySerGlnLeuProThrCyGlyAlaProGlnValYrSerValaArgArg 600
DB 2327 CTGGCTTACAGCTGACACGACGCTGCCCATCTTGCCGAGAACCAAGCTCTATTCTGTGCGCAGGC 2386
QY 600 rGlnLysProSerAspArgAlaAspSerArgArgSerTrpHisGlnGlySerProPhg 620
DB 2387 GGCAGAGCCCAAGTGAAGAGCTGACTCGCGGCGAGCTGGCATGTGAAGAGAGCCCTTTGG 2446
QY 620 lylsGlnPheylsArgArgSerCyGlnMerGlnPheGlnGlySerIleMetSerGlnA 640
DB 2447 AAAGCAGTTTAAACGACAGAGCTGCCAATGGAATTTGGAGAGACATCATCTCAGAGA 2506
QY 640 snArgSerArgGlnGlyLeuValGlySerGlnSerSerPheSerGlySerMetG 660
DB 2507 ACAGGTACAGCGGAGAGCTGGGAGAAAGTGGCAGTCACTAGCTTTTGGGCGCAGCATGG 2566
QY 660 lullellegluValSer 665
DB 2567 AAATCATTAAGGCTCC 2583
RESULT 19
ABV21312
ID ABV21312 standard; cDNA; 5145 BP.
XX ABV21312;
AC
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 21303.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN MO200160860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-018319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
XX
PS Claim 1; Page 3539; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (1) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification of its complement. (1) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the efficacy
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)

CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (1) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
SQ Sequence 5145 BP, 1408 A; 1135 C; 1253 G; 1346 T; 0 U; 3 Other;
Alignment Scores:
Pred. No.: 0 Length: 5145
Score: 472.00 Matches: 663
Percent Similarity: 99.40% Conservative: 0
Best Local Similarity: 99.40% Mismatches: 2
Query Match: 70.98% Indels: 4
DB: 5 Gaps: 0
US-10-029-345a-109 (1-665) x ABV21312 (1-5145)
QY 1 MetAlaHisGluMetIleGlyThrGlnIleValThrGluArgLeuValAlaLeuLeuGln 20
DB 589 ATGGCCCATGAGATGATTGGAATCAATTTGTAATGAGAGGTGGCTGTGAGAA 648
QY 21 SerGlyThrGluValLeuLeuIleAspSerArgProPheValGlnTrpAsnThrSer 40
DB 649 AGTGAAACGAAAAAGTGTGCTTAATGATGACCCGCGCATTTGTGAAATACAAATACATCC 708
QY 41 HisIleLeuGluAlaIleAsnIleAsnCySerIleValMetIleArgArgLeuGlnGln 60
DB 709 CACATTTTGAAGCCATTATATCAATCTCTCAAGCTTATGAGGAAGTTGGCAACAG 768
QY 61 AspLysValIleuIleThrGluLeuIleGlnHisSerAlaLysHisLysValAspIleAsp 80
DB 769 GACAAAGTGTATTAACAGAGCTCATCCAGATTCAAGCAAACTATGAAGTTGACATTGAT 828
QY 81 CySerGlnLysValValValTyrAspGlnSerSerGlnAspValAlaSerLeuSerSer 100
DB 829 TCCAGTCAGAAAGTTGTATTTACATCAATCAAACTCCCAAGATGTGGCTCTCTCTTCA 888
QY 101 AspCyPheLeuThrValLeuLeuGlyLysLeuGlnLysSerPheAsnSerValHisLeu 120
DB 889 GACTGTTTCTACGTCTACTTCTGGTAACTCGAGAAAGCTTCAACTCTGTCCACTCG 948
QY 121 LeuAlaGlyGlyPheAlaGluPheSerArgCyPheProGlyLeuCyGlnGlyLysSer 140
DB 949 CTTCAGAGTGGGTTGTGCTAGTTCTCTCGTTTCCCTCGGCTCTGTGAAGAAATCC 1008
QY 141 ThrLeuValProThrCyGlnSerGlnProCyLeuProValAlaAsnIleGlyProThr 160
DB 1009 ACTCTAGTCCCTACCTGCACTTCTCAGCCTTCTTACCTGTGGCCAACTTGGGCCAACCC 1068
QY 161 ArgIleLeuProAsnLeuTyrLeuGlyCyGlnArgAspValLeuAsnLysGlnLeuIle 180
DB 1069 CGAATTTCTTCCAAATCTTTATCTTGCTGCCAGCGAGATGCTCTCAACAAAGAGCTGAT 1127
QY 181 -GlnGlnAsnGlyIleGlyTyrValIleAsnAlaSerTyr-ThrCySProLysProAsp 200
DB 1128 GCAGCAGATGGGATGTTGTTATGTGTTAAATGCCAGAA-TACGTGTCCAAAGCCGACT 1186
QY 200 heIlePProGlnSerHisPheLeuArgValProValAsnAspSerPheCyGlnLysIle 220
DB 1187 TTATCCCCAGCTCATTTCTCGCGTGCCTGTGAATACAGCTTTGTGAAATTT 1246
QY 220 euProTrpLeuAspLysSerValAspPheIleGlnLysAlaLysAlaSerAsnGlyCyArg 240
DB 1247 TGCCTGTGTGACAAATCATGTAATTTCTATGAGAAACAAAGCTCTCAATGATGTG 1306
QY 240 alleuValHisCySerLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIle 260
DB 1307 TTCTAGTGACAGTTTAAAGTGGGATCTCCCGCTCCGACCACTCGCTATCGCTACATCA 1366
QY 260 eLysArgMetAspMetSerLeuAspGluAlaTyrArgPheValLysGlnLysArgPro 280
DB 1367 TGAAGAGGTGACATGCTTTAGATGAAGCTTACAGATTTGTGAAGAAAAAGACCTTA 1426
QY 280 hrlIleSerProAsnPheAsnPheLeuGlyGlnLeuLeuAspTyrGlnLysLysIleLysA 300

```
Db 1427 CTATATCTCCAACTTCAATTTCTGGGCCAATCTCCGACTATGAGAGAAATTTAAGA 1486
Qy 300 snglnthrGlyAlaSerGlyProLysSerLysLysLeuLysLeuLysLeuLysPro 320
Db 1487 ACCAGACTGAGAGATCAGGGCCAAAGCAAACTCAAGCTGCTGCACTTGAGAAACCA 1546
Qy 320 snglnProValProAlaValSerGluGlyGlyGlyLysSerGluThrProLysSerPro 340
Db 1547 ATGAACTGCTCCCTGCTGCTCTCAGAGGGTGAAGAAAGCCGAGAGCCCTCAAGTCCAC 1606
Qy 340 roCySalaaPSSerAlaThrSerGluAlaAlaGlylnarProValHisProAlaSerY 360
Db 1607 CTTGTGCCGACTCTGCTACTCTCAGAGGACAGAGCAAAAGCCCGTGCATCCGCCACAGG 1666
Qy 360 aLProSerValProSerValGlnProSerLeuLeuGluAspSerProLeuValGlnAla 380
Db 1667 TGCCAGCCGTCAGGAGGTGCACCGCTCGCTTAGAGGACAGCCCGCTGTGACAGGCC 1726
Qy 380 euSerGlyLeuHisLeuSerAlaAspArgLeuGluAspSerLeuLysLeuLysArgSer 400
Db 1727 TCAGTGGGCTGCACCTGTCCTCCGACAGAGGCTGGAGAGCAAGCAATTAAGCTCAAGCCT 1786
Qy 400 heSerLeuAspLysSerValSerTyrSerAlaSerMetAlaAlaSerLeuHisGlyP 420
Db 1787 TCTCTCTGATATCAATCAGATTTCATATTACAGCCAGCATGGCAGCATCTTACATGCT 1846
Qy 420 heSerSerSerGluAspAlaLeuGluTyrTyrLysProSerThrThrLeuAspGlyTh 440
Db 1847 TCTCTCATCAGAAAGATGCTTGGAATACTACAAACTTCCACTCTGATGGGACCA 1906
Qy 440 snLysLeuCySGlnPheSerProValGlnGlyLeuSerGluGlnThrProGluThrSerP 460
Db 1907 ACAAGCTATGCCAGTCTTCCCTCGTTCAAGAACTATCGAGCAGACTCCCGAAACCA 1966
Qy 460 roAspLysGluGluAlaSerTyrProLysLysLeuGlnThrAlaArgProSerAspSer 480
Db 1967 CTGATTAAGAGAGAACCCACATCCCAAGAAAGCTGACAGCCGACGCTTACAGACAG 2026
Qy 480 lAsSerLysArgLeuHisSerValArgThrSerSerSerTyrThrAlaGlnArgSerLeu 500
Db 2027 AAGAGCAAGCATGTCCTGCTGTCAGAAACAGAGCATGTCACCGCCAGAGAGTCCCTT 2086
Qy 500 euSerProLeuHisArgSerGlySerValGluAspAsnTyrHisThrSerPheLeuPhe 520
Db 2087 TATCTCCACTGCATGCAAGTGGAGAGGTGGAGAGCAATTAACAACACACTTCTTTG 2146
Qy 520 lYLeuSerThrSerGlnGlnHisLeuThrLysSerAlaGlyLeuGlyLeuLysGlyTyrP 540
Db 2147 GCTTTCCACAGCCAGCAGACACTCAAGAGTCTGCTGGCTTGAAGGGCTGGC 2206
Qy 540 lAsSerAspLysLeuAlaProGlnThrSerThrProSerLeuThrSerSerTyrPheAs 560
Db 2207 ACTCGGATATCTTGGCCCCCAGACTCTAACCCTTCCCTGACAGAGCTGATTTTGG 2266
Qy 560 lAThrGlnSerSerHisPheTyrSerAlaSerAlaIleTyrGlyLysSerAlaSerTyrS 580
Db 2267 CCACAGAGCTCCACACACTTCTACTGCTGCTCAGCCATCTACGAGAGGACGACTTACT 2326
Qy 580 eAlaTyrSerCySGlnLeuProThrCySGlyAspGlnValTyrSerValArgArg 600
Db 2327 CTGCCTACAGCTGACAGCCAGCTGCCCACTTGGCGAGACCAAGTCTATCTGCGCAG 2386
Qy 600 rGglnLysProSerAspArgAlaAspSerArgArgSerThrHisGlnLysSerProPhe 620
Db 2387 GCGAAGAGCAAGTACAGAGTGACTGCGGGCGAGCTGGGATGAAGAGAGGCCCTTTG 2446
Qy 620 lLysGlnPheLysArgArgSerCySGlnMetGlnPheGlyGlnSerLysMetSerGln 640
Db 2447 AAAAGCAGTTTAAAGCAGAGAGCTGCCAAATGGAATTTGGAGAGAGCATCATGCA 2506
Qy 640 snArgSerArgGluGluLeuGlyLysValGlySerGlnSerSerPheSerGlySerMetG 660
```

```
Db 2507 ACAGGTCACGGGAGAGAGCTGGGGAAGTGGGCACTGACTTACGCTTTCCGGCAGCAT 2566
Qy 660 lLillelGluValSer 665
Db 2567 AAATCATTAAGGCTCTCC 2583
RESULT 20
ABV21316
ID ABV21316 standard; cDNA, 5145 BP.
XX
AC ABV21316;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 21307.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO200160860-A2.
XX
PD 23-AUG-2001.
XX
PE 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUN-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
PS Claim 1; Page 3540; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the efficacy
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (I) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
SQ Sequence 5145 BP; 1408 A; 1135 C; 1253 G; 1346 T; 0 U; 3 Other;
Alignment Scores:
Pred. No.: 0 Length: 5145
Score: 472.00 Matches: 663
Percent Similarity: 99.40% Conservative: 0
Best Local Similarity: 99.40% Mismatches: 2
Query Match: 70.98% Indels: 4
DB: 5 Gaps: 0
US-10-029-345A-109 (1-665) x ABV21316 (1-5145)
Qy 1 MetAlaHisGluMetIleGlyThrGlnIleValThrGluArgLeuValAlaLeuLeuGlu 20
```

Db 589 ATGGCCCATGATGATGGAACTCAATTGTTACTGAGAGTGGGCTGCTGAGAA 648
 QY 21 SerGIYThGluYValleuLeuIIeaSPSerArgProPheValGluYrhanThSer 40
 Db 649 AGTGAACGGAAAAAGGCTGCTAAATTAGAGCCGGCCATTGTGGAAATPACAATACATCC 708
 QY 41 HsAlleuGluIuaIleauNIleaenCySerIleuMeLysArgArgLeuGlnGln 60
 Db 709 CACATTTTGGAGCCATTATATCACTGCTCCAACTTATGAAAGGAGTTGGCAACAG 768
 QY 61 AspIleValleuIleThrGluLeuIIeGlnHsSerAlaYshIleYValAspIleAsp 80
 Db 769 GACAAAGTGTAAATTACAGGCTCATCCAGCATTCAGCCAAACATAGAGTTGACATTGAT 828
 QY 81 CySerGIleYValValIValIYrAspGInserSerGIAspValIAserLeuSerSer 100
 Db 829 TGGAGTACGAAGGTGTAGTTAGATTCAGATCAAAAGCTCCCAAGATGTCCCTCTCTCTTA 888
 QY 101 AspCyPheLeuThrValIleuLeuGIlyIleuGIlyIleuSerPheAsnSerValHsIleu 120
 Db 889 GACTGTTTCTCATCTGACTTCTGGGTAAACTGGAAAGAGCTTCAACTCTGTTCACTCG 948
 QY 121 LeuAlGIlyGlyPheAlaGluPheSerArgCyPheProGIlyLeuCyAspGluGIlySer 140
 Db 949 CTTCAGAGTGGGTTTCTGCTGAGTTCTCTCGTTGTTCTCGCTGCTGTGAAGAAATCC 1008
 QY 141 ThrLeuValProThrCyValIleSerGInProCyIleuProValAlaAsnIleGIlyProThr 160
 Db 1009 ACTCTAGTCCCTTACCTGCAATTTCTCAGCCTTGTGCTTACCTGTGCAACATGGGCCAAC 1068
 QY 161 ArgIleLeuProAsnLeuThrLeuGIlyCyGlnArgAspValIleuAsnIleuIle 180
 Db 1069 CCAATTCCTCCCAATCTTTATCTCTGCTGCCAGCCAGATGCTCTCAACAGAGCTGAT- 1127
 QY 181 -GlnGlnaangIyIleGIlyYrValleuAsnIleSerIYr-ThrCyProIYsProAsp 200
 Db 1128 GCAGCAGATGGGATTGATTATGTATTAAATGCCACAA-TACTCTCCAAAGCCGACT 1186
 QY 200 HeIleProGluSerHsPheLeuArgValProValAsnAspSerPheCyGluYValIle 220
 Db 1187 TTAATCCCGAGTCTCATTTCTCGCGTGTGCTGTGATGATGACAGCTTTTGTGAAATTT 1246
 QY 220 euProTrIleuAspIYsSerValIAspPheIIeGIlyValIleYValIAserAngIYCyAs 240
 Db 1247 TCCCGGTGTGGACAAATCAGTATGATTTCAATGAGAAAGCAAAAGCTCCAAATGATGAG 1306
 QY 240 alleuValHsCyLeuAlaGIlyIleSerArgSerAlaThrIleAlaIleAlaYrIle 260
 Db 1307 TTTAGTGTGACTGTATTAGCTGGGATCTCCGGCTCCGCCACATCGCTATCGCTACATCA 1366
 QY 260 eLlyeArgMetAspMetSerIleuAspGluAlaIYrArgPheValIYsGIlyYValArgPro 280
 Db 1367 TGAAGAGATGACATGCTTTAGATGAAGCTTACGATTTGTGAAGAAAAAGCCCTTA 1426
 QY 280 hrIleSerProAsnPheAsnPheLeuGIlyIleuLeuAspIYrGIlyYValIleYVal 300
 Db 1427 CTRATCTCCAACTTCAATTTCTTGCGCAACTCTCGACATGTGAGAAAGATTTAA 1486
 QY 300 angIlnThrGIlyIAserGIYProIYsSerIleuYValleuLeuHsIleuGIlyYPro 320
 Db 1487 ACCAGACTGGAGCATCAGGGCCAAAGAGCAACTCAAGTGTGCTGCACCTGGAGAGCCAA 1546
 QY 320 angIuProValProAlaValIserGIlyGIlyIYsSerGIuThrProLeuSerPro 340
 Db 1547 ATGAACCTGTCCCTGTCTCTCAGAGGTGACAGAAAGCCAGACGCCCCCTCAGTCCAC 1606
 QY 340 roCyAlaAspSerAlaThrSerGIuAlaAlaGIlyIYsIlnProValHsProAlaSerV 360
 Db 1607 CCTGTGCCCACTCTGCTACCTCAGAGGCGACGAGCAAAAGCCCGGTGATCCCGCCAGG 1666
 QY 360 aIProSerValIProSerValIlnProSerLeuLeuGIuAspSerProLeuValIlnAl 380
 Db 1667 TGCCTGAGCTGCGCCAGCGTGCAGCGCTGCTGTATGAGACACGCCGCTGTATACAGGCC 1726

QY 380 euSerGIYleuHsIleuSerAlaAspArgLeuGIuAspSerAnIYsLeuYValArgSerP 400
 Db 1727 TCAGTGGGCTGACCTGTCCGACAGAGCTGGAAGACAGCAATAGCTCAAGCGTTCC 1786
 QY 400 heSerLeuAspIleYsSerValIserIYsSerAlaSerMetAlaIAserLeuHsGIlyP 420
 Db 1787 TCTCTGTGATATCAATCAGTTTCAATATTCAGCCAGCATGGCAGCATCTTACATGSGCT 1846
 QY 420 heSerSerSerGIuAspAlaLeuGIlyYrIYsProSerThrThrLeuAspGIlyThA 440
 Db 1847 TCTCTCATCAGAAAGATGCTTTGGAATCTAAGAACTTCCACTACTGTGATGGGACCA 1906
 QY 440 enIYleuCyGlnHsSerProValIlnGlnIleuSerGIuIlnThrProGIuThrSerP 460
 Db 1907 ACAAGCTATGACAGATCTCCCTGTTCAGAACTATGGAGCAAGCTCCGAAACCAAGTC 1966
 QY 460 roAspIleGIlyIuIAserIleProIYsIleuGIlnThrAlaArgProSerAspSerG 480
 Db 1967 CTGATAGAGAGAACCCAGCATCCCAAGAGCTCAGACCCCGAGCTTACAGACGCC 2026
 QY 480 IlnSerIYsArgLeuHsIserValArgThrSerSerSerGIYThrAlaGlnArgSerLeu 500
 Db 2027 AGAGCAAGCGATTGATTCGATCAGAACACAGACAGATGGCACCCCGAGAGTCCCTTT 2086
 QY 500 euSerProLeuHsIArgSerGIYsSerValGIuAspAsnThrHsThrSerPheLeuPheG 520
 Db 2087 TATCTTCATGTGATGAAGTGGAGCGTGGAGAGCAATTAACACACAGCTTCTTTTTCG 2146
 QY 520 IYleuSerThrSerGlnGlnIleuThrIleuThrIYsSerAlaGIlyleuGIlyYValIYrP 540
 Db 2147 GCCTTCCACCGACCGACAGCACTCAGCAAGTCTGCTGGGCTTAAAGGCTGGC 2206
 QY 540 IAserAspIleuAlaProGlnThrSerThrProSerLeuThrSerSerIYrYrPheA 560
 Db 2207 ACTCGAATATCTTGCCCCCGCCAGACTCTAACCCCTTCCCTGACAGAGCTGTATTTCG 2266
 QY 560 IATHGluSerSerHsPheYrSerAlaIserAlaIleYrGIlyIleSerAlaSerIYs 580
 Db 2267 CCACAGAGCTCCACACTTCTACTGCTGCTGACCATTCAGAGAGCAGTGCAGTACT 2326
 QY 580 eAlaIYsSerCySerGlnIleuProThrCyGIYAspGlnValIYrSerValIArgArgA 600
 Db 2327 CTGCTTACAGCTGCGACCGACCTGCCACTTGGGAGAACCAAGTCTAATTTCTGCGCAGCC 2386
 QY 600 IYGIlyIleProSerAspArgAlaAspSerArgArgSerIYrHsGIlyIleuSerProPheG 620
 Db 2387 GGCAGAGCCAAAGTACAGAGCTGACTCGCGCGAGCTGGCATGAAGAGACCCCTTTG 2446
 QY 620 IuYsGlnPheIYsArgArgSerCyGlnMetGluPheGIlyIleuSerIleMetSerGIuA 640
 Db 2447 AAAGCAGATTTAAGCAGAAAGCTGCCAAATGGAATTTGAGAGAGCATCATGTACAGAA 2506
 QY 640 snArgSerArgGIlyIleuGIlyYsValGIYsSerGIYsSerSerPheSerGIYsSerMetG 660
 Db 2507 ACAAGTCAAGGAAAGCTGGGGAAAGTGGGAGTCAAGTCACTTTTGGGAGAGATGG 2566
 QY 660 IuIleIleGIlyValIser 665
 Db 2567 AAATCATTAGAGTCTCC 2583

RESULT 21
 ABV26826
 ID ABV26826 standard; cDNA, 5145 BP.
 XX
 AC ABV26826;
 XX
 DT 16-SEP-2002 (first entry)
 XX
 DE Human prostate expression marker cDNA 26817.
 XX
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; ss.

XX	Homo sapiens.
OS	WO200160860-A2.
PN	23-Aug-2001.
PD	20-FEB-2001; 2001WO-US005171.
XX	17-FEB-2000; 2000US-0183319P.
PR	16-MAR-2000; 2000US-0189862P.
PR	25-MAY-2000; 2000US-0207454P.
PR	09-JUN-2000; 2000US-0211314P.
PR	18-JUL-2000; 2000US-0219007P.
PR	13-DEC-2000; 2000US-0235281P.
XX	(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PA	Schlegel R, Endege WO, Monahan JE;
P1	WPI: 2001-662795/76.
DR	
XX	
XX	Novel isolated nucleic acid molecule associated with cancerous state of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer.
PT	
PS	Claim 1; Page 5420-5421; 11750pp; English.
CC	The invention relates to an isolated nucleic acid molecule (1) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (1) is useful for: (a) assessing whether a patient is afflicted with prostate cancer; (b) monitoring the progression of prostate cancer in a patient; (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient; (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient; (e) selecting a composition for inhibiting prostate cancer in a patient; (f) assessing the prostate cell carcinogenic potential of a compound; (g) determining whether prostate cancer has metastasized in a patient; (h) assessing the aggressiveness or indolence of prostate cancer in a patient ; (1) is also useful as a pharmacodynamic or pharmacogenomic marker
CC	
CC	
SO	Sequence 5145 BP; 1408 A; 1135 C; 1253 G; 1346 T; 0 U; 3 Other;
Alignment Scores:	
Pred. No. :	0 Length: 5145
Score:	472.00 Matches: 663
Percent Similarity:	99.40% Conservative: 0
Best Local Similarity:	Mismatches: 2
Query Match:	Indels: 4
DB:	Gaps: 0
US-10-029-345A-109 (1-665) x ABV26826 (1-5145)	
OY	1 MetAlAHISglMeTlIEglYThrgInIlleValThrGlUArgLeuValAlaleuleuglu 20
Dd	589 ATGGGCCATGAGATGATTGAACCTCAAAATTGTACTGAGAGGTTGGTGCTGCCTGGA 648
OY	21 SerGIlyThrGIuLyVaIleuLeuIlleaSpSerArGProPhyeValGIuTyraAntHrSer 40
Dd	649 AGTGAAACCGAAAAAGTCGTCTTAATVGATTAGCCGGCATTGTGTGGAATAACAATAATCC 708
OY	41 HisIIeLeuGlnAlAleasnIlleaSnCySserLYLeMeTrysArGrAgLeuGlnGln 60
Dd	709 CACATTTTGGAGGCCATTAAATCATCTGCTCCAACCTTATGAAAGCAAAGTTGCAACAG 768
OY	61 AsPlYsValIleuIlleThrGlueuIlleGlnHisSerAlALyshiSLySValAspiIleasp 80
Dd	769 GACAAAGTCTTAATTACAGAGCCATCCAGCATTCAGCCAAAATAGAAGTTGACATTGAT 828
OY	81 CySSerGIuLyVaIVaIlyValryApGIInsErSerGIuAspValAlaserLeuSerSer 100
Dd	829 TCAGTCAGAAAGTTTAGATTACGATCAAAAGCTCCCAAGATGTTGCCTCTCTCTTCA 888

QY	101	AspCysPheLeuThrValLeuLeuGluYlyrSerGluYlyrSerPheAsnSerValHisLeu	120
Db	889	GACGTGTTTCCACCTGTACTTCTGGGTAAACCTGGAAGAGCTTAACCTGTTACCGG	948
QY	121	LeuAlaGlyGlyPheAlaGluPheSerArgCysPheProGlyLeuCysGluGluYlyrSer	140
Db	949	CTTGACAGGTGGGTTTGCTGAGTTCTCTCGTTGTTTCCCTGGGCTCTGTGAAGAAATCC	1008
QY	141	ThrLeuValProThrCysIleSerGlnProCysLeuProValAlaAsnIleGlyProThr	160
Db	1009	ACTGACGACCTACCTGCATTTCTCAAGCCTTGCTTACCTGTTGCCAACATTGGGCAACC	1066
QY	161	ArgIleLeuProAsnLeuTyrLeuGlyCysGlnArgPheValLeuAsnIleGluLeuIle	180
Db	1069	CGAATTTCTCCCACTTTATCTTGGCTGCCAGCAAGATGTCTCAACAAGAGCTGAT-	112
QY	181	-GlnGlnAsnGlyIleGlyTyrValLeuAsnAlaSerTyr-ThrCysProIysProAspP	200
Db	1128	GCAGACAGATGGGATGGTTATGTGTAAATGCCAGCAA-TACCTGTCCAAAGCCTGACT	1180
QY	200	helleProGluSerHisPheLeuArgValProValAsnAspSerPheCysGluYlyrIle	220
Db	1187	TTATCCCCGAGCTCATTTCTCGGTGGCCGTGAAGACACTTTGTGTGAATAATT	124
QY	220	eupProTyrLeuAspIysSerValAspPheIleGluYlyAlaIysAlaSerAsnGlyCysV	240
Db	1247	TGCGGTGGTTGACAAATCAGTAGATTTCATTGGAAGAAACAAACCTCCAAATGATGTG	1300
QY	240	AlleuValHisCysLeuAlaGlyIleSerArgSerHisThrIleAlaIleAlaTyrIleM	260
Db	1307	TTCTAGTGACATGTTTACTGTGGACTCCGCTCCGCAACATCGCTATCGCTACATCA	1366
QY	260	etIysArgMetAspMetSerLeuAspGluAlaTyrArgPheValIysGluYlyArgProT	280
Db	1367	TGAAGAGGATGGACATGTCTTAGTAGAGCTTACAGTTGTGAAAGAAAAGAACCTTA	1422
QY	280	hrlIleSerProAsnPheAsnPheLeuGlyGlnLeuLeuAspTyrGluYlyIleYlysa	300
Db	1427	CTATATCTCCAAACTTCATTTTCTGGGCCCAACTCTGACATATAGAAAGAGATTAACA	1488
QY	300	snGlnThrGlyAlaSerGlyProIysSerIysLeuYlyLeuLeuHisLeuGluYlyProC	320
Db	1487	ACCAGACTGGAGCATCAGGCGCCAAAGACAAACTCAAGCTGTGACCTGGAGAAAGCCA	1544
QY	320	snGluProValProAlaValSerGluGlyGlnYlySerGlnThrProLeuSerProp	340
Db	1547	ATGAACCTGTCCCTGCTGTCTCAGAGGGGTGGACAAAGAAAGCAGCGCCCTCACTCCAC	1600
QY	340	roCysAlaAspSerLamPheSerGluAlaAlaGlnArgProAlaHisProAlaSerY	360
Db	1607	CCTGGCCGAGCTGTCACTCACTGAGGGACAGGACAAAGGCCCTGTCAATCCGCCAGCG	1666
QY	360	alProSerValProSerValGlnProSerLeuLeuGluAspSerProLeuValGlnAlaL	380
Db	1667	TGCCAGCGTGGCCAGCGGACGCCGTGCTTAGAGACAGCCGCTGTGATACAGCGCC	1722
QY	380	eusErgIyLeuHisLeuSerAlaAspArgLeuGluAspSerAsnYlyLeuYlyAspSerP	400
Db	1727	TCAGGGGCTGACCTGTCTCCGACAGAGCTGGAGAACAGCATATAGCTCAAGCGTTCT	1788
QY	400	hesELeuAspIleYlySerValSerTyrSerAlaSerMetAlaIaSerLeuHisGlyP	420
Db	1787	TCTCTCTGATATCAATCAAGTTTATATTACGCCAGCATGCGACATCTTAACTAGGCT	1844
QY	420	hesEserSerErgIuAspAlaLeuGluTyrTyrIlyProSerThrThrLeuAspGlyThrA	440
Db	1847	TCTCTCTATCAGAAATGCTTTGGAAATCTCAAAACCTTCCACTCTCGATGGGACCA	1900
QY	440	snIlyLeuCysGlnPheSerProValGlnIleLeuSerGluGlnThrProGluThrSerP	460
Db	1907	ACAACTATGCGAGTTCTCCCTGTTCAGAACTATCGAGACAGACTCCGAAACCAAGTC	1966
QY	460	roAspIyGluGluAlaSerIleProYlyIyLeuGlnThrAlaArgProSerAspSerG	480

```
DB 1967 CTGATAGAGGAGGAGCAGATCCCAAGAGCTGAGACCGCCAGCCTTCAGACAGCC 2026
QY 480 InSerIyArGLeuHhISeSerValArGThSeSerSerGlyThAlaGlnArGSeLeuL 500
DB 2027 AGAGCAGAGGATGCTTCGCTCAGAACAGCAGCGAGTGGCACCGCCAGAGGTCCTTT 2086
QY 500 euseProLeuHhISeSerGlySeValGlnuApAnTyThISeSerPheLeuPhG 520
DB 2087 TATCTTCAGCTGCAATCCAGATGGAGCGTGAGGACATTAACACACAGCTTCCTTTTG 2146
QY 520 IyLeuSerThSeSerGlnGlnHhISeLeuThLySeSerAlaGlyLeuGlyLeuGlyTrpH 540
DB 2147 GCCTTCCACAGCCAGCAGACACCTACAGAGCTCTGCGCCCTTAAGCGCTGGC 2206
QY 540 ISeSerAplleuAlaProGlnThSeSerThProSeLeuThSeSerTyrPheA 560
DB 2207 ACTCGATATCTTGCGCCCGCCAGACCTTAACCTTCCTCCAGACAGCAGCTGGTATTTG 2266
QY 560 IaThGlnuSeSerHhISePheTySeSerAlaSerAlaIeTyGlyGlySeSerAlaSeTyS 580
DB 2267 CCACAGATCTTACACTTCTACTGCTGCTCAGCCCATTAAGGAGGAGTGCAGTACT 2326
QY 580 eAlaIeTySeSerCySeSerGlnLeuProThCySeGlyASePlnValTySeSerValArGArG 600
DB 2327 CTGCCACAGCTGCAGCCAGCTGCCCATCTGGGAGACCAAGCTATTTCTGCGCAGGC 2386
QY 600 rGlnLyArProSeArPaArGAlaASeSerArGArGSeTyrHhISeGlnuSeSerProPhG 620
DB 2387 GCGAGAGGCAAGTGCACAGAGCTGACTCGCGGCGAGCTGGGCTGAAGAAGAGCCCTTG 2446
QY 620 IuLySeGlnPheLyArGArGSeSerCySeGlnMeGlnuHhGlyGlnuSerIleMeSeSerGln 640
DB 2447 AAAAGAGATTAAACGACAGAGCTGCCAATGGAAATTTGAGAGAGACCATCATCTCGAGA 2506
QY 640 sArGSeArGArGlnuLeuGlyLySeValGlySeSerGlnuSeSerPheSeSerGlySeMeG 660
DB 2507 ACAGGCTCAGCGAGAGAGCTGGGAAAGTGGGAGTACGCTTACCTTTTGGGACAGATG 2566
QY 660 IuIleIleGlnuAlSeSer 665
DB 2567 AAATCATGTAGGCTCC 2583

RESULT 22
ABV27131
ID ABV27131 standard; cDNA; 5145 BP.
XX
XX ABV27131;
AC 16-SEP-2002 (first entry)
XX
DT Human prostate expression marker cDNA 27122.
XX
DE Human prostate expression marker cDNA 27122.
XX
KM Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX pharmacogenomic marker; gene; ss.
XX Homo sapiens.
XX OS
XX WO200160860-A2.
XX PN
XX 23-AUG-2001.
XX
XX 20-FEB-2001; 2001WO-US005171.
XX PF
XX 17-FEB-2000; 2000US-0183119P.
XX PR 16-MAR-2000; 2000US-0189862P.
XX PR 25-MAY-2000; 2000US-0207454P.
XX PR 09-JUN-2000; 2000US-0211314P.
XX PR 18-JUL-2000; 2000US-0219007P.
XX PR 13-DEC-2000; 2000US-0255281P.
XX
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PA
```

```
PI Schlegel R, Endege WO, Monahan JE;
XX
XX WPI; 2001-662795/76.
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
XX Claim 1, Page 5499-5500; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (1) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV0010-ABV62213) of the
CC specification or its complement. (1) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (c) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (1) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
SQ Sequence 5145 BP; 1408 A; 1135 C; 1253 G; 1346 T; 0 U; 3 Other;

Alignment Scores:
Pred. No.: 0 Length: 5145
Score: 472.00 Matches: 663
Percent Similarity: 99.40% Conservative: 0
Best Local Similarity: 99.40% Mismatches: 2
Query Match: 70.98% Indels: 4
DB: 5 Gaps: 0

US-10-029-345A-109 (1-665) x ABV27131 (1-5145)
QY 1 MetAlaHhISeGMeTIlleGlyThrGlnIleValJThrGluArGLeuValAlaLeuLeuGln 20
DB 589 ATGGCCCATGAGATGATGAACTGAATTTGTTACTGAGAGGTTGGCTCTGCTGGAA 648
QY 21 SerGlyThrGlnuLySeValleuLeuIleASeSerArGProPhValGlnuTyArAnThSeSer 40
DB 649 AGTGGAGCGGAAAGAGTGTCTTAATGTATACCCGGCCATTTGTGGAAATACATATACATCC 708
QY 41 HhIleLeuGlnuAlaIleAhnIleAhnCySeSerISeuMeTyArGArGLeuGlnGln 60
DB 709 CACATTTTGGAGCCATTAATATCAATGCTCCAGCTTATGAAGGAGGTTGCAACAG 768
QY 61 ASePlySeValleuIleThGlnuLeuIleGlnHhISeSerAlaIeHhISeValASePleASeP 80
DB 769 GACAAAGTGTATTAACAGAGCTCATCCAGCATTCAGCCAAACATTAAGGTTGACATTGAT 828
QY 81 CySeSerGlnuLySeValValValTyArASePInSeSerGlnuASePValAlaSeSerLeuSeSer 100
DB 829 TGCAGTCCAGAAAGTTGTAGTTTACGTAAAGCTCCCAAGAGTTGGCTCTCTCTTCA 888
QY 101 ASePlySePheLeuThValleuLeuGlyLySeLeuGlnuLySeSerPheAhnSeSerValHhISe 120
DB 889 GACTGTTTTCTCAGCTACTCTCTGGGTAAACTGAGAGAGAGCTTCACACTCTGTTACCTG 948
QY 121 LeuAlaGlyGlyPheAlaGlnuPheSeSerArGlyPheProGlyLyLeuCySeGlnuGlyLySeSer 140
DB 949 CTTCAGAGGAGGTTGCTGAGTTCCTCTGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1008
QY 141 ThrLeuValProThCySeIleSeSerGlnProCySeuProValAlaAhnIleGlyProTh 160
DB 1009 ACTCTAGTCCCACTGATTTCTCAGGCTTGCTTACCTGTTGCCAACAATGGGCAAC 1068
QY 161 ArgIleLeuProAhnLeuTyTrleuGlyCySeGlnuArGArGValleuAhnLySeGlnuLeu 180
DB 1069 CGAATTCCTCCCAATCTTATCTTGGCTGCCAGCAGATGCTCTCAACAAGAGAGCTGAT- 1127
QY 181 -GlnGlnAhnGlyIleGlyTyTrValleuAhnAlaSeSerTyTr-ThCyASePlyProASeP 200
```

Db 1128 GCACGAGATGGATGGTTATGTTAAATGCCAGCA--TACCTGTCCAAGCTTGACT 1186
Qy 200 heileProgluSerHisPheleuabrgValProValaAanapSerPheCysGluys11el 220
Db 1187 TTATCCCCGAGTCTCATTTCTCGGTGTGCTGGAATGACACTTTTGTGAAGAAAATTT 1246
Qy 220 eufProTIPLeuAspLysSerValaAspPhe11egluysAlaLysAlaSerAsnG1yCysv 240
Db 1247 TGCCGTGGTTGGACAATCATGATGATTTCATTGAGAAAGCAAAAGCCTCCAAATGATGTG 1306
Qy 240 alleuValHisCysleuAlaG1y11eseraYserAlaThr11eAla11eAlaTyr11ew 260
Db 1307 TTTAGAGCACTGTTTAGCTGGAGATCTCCGCTCCGCAACATCGCTATCGCTTACATCA 1366
Qy 260 etLysaTgMetAspMetSerLeuAspGluAlaTyrAArgPheValLysGluLysAArgProT 280
Db 1367 TGAAGAGATGAGCAATCTTTTGAATGAGACTTACGATTTTGAAAGAAAAGACCTTA 1426
Qy 280 hr11eserProaAnPheAsnPheleuG1yG1nleuAspTyrG1uLys11eLysA 300
Db 1427 CTATATCTCCAAACTTCAATTTTCTGGGCCAACTCCGAGCATATGAGAAAGATTTAAGA 1486
Qy 300 snG1nThrG1yAlaSerG1yProLysSerLysleuLysleuHisleuG1uLysProA 320
Db 1487 ACCAGACTGGAGATCAGAGGCGCAAGAGCAAACTCAAGCTGTGCACTTGAGAGAGCAA 1546
Qy 320 snG1uProValProAlaValSerG1uG1yG1nLysSerG1uThrProLysSerProP 340
Db 1547 ATGAACTGTCTCTGTCTGTCTCAGAGGCTGAGCAAGAAAGCGAGAGGCCCCCTCAGTCCAC 1606
Qy 340 roCySa1aAspSerAlaThrSerG1uAlaAlaG1yG1nAArgProValHisProAlaSerV 360
Db 1607 CCTGTGCCGACTGTGCTACCTCAGAGGCAAGCAAAAGCCGCGATCCCGGCAAGG 1666
Qy 360 a1ProSerValProSerValG1nProSerleuLysleuAspSerProLeuValG1nAla 380
Db 1667 TGCCCAAGCTGGCCAGAGTGCGACCGTGTGTGAAGACAGCCGCTGTGACAGGCGCC 1726
Qy 380 eusSerG1yleuHisleuSerAlaAspAArgLeuG1uAspSerAsnLysleuLysAArgSerP 400
Db 1727 TCAGTGGGCTGAGCCGTGTGCGAGACAGGCTGGAAGACAGCAATAGCTCAAGCGTTCTT 1786
Qy 400 heSerleuAsp11elysSerValSerTyrSerAlaSerMetAla1aSerleuHisG1yP 420
Db 1787 TCTCTCTGATATCAAAATCAGTTTCATATTCAGCCAGCATGCGACATCTTACATGCT 1846
Qy 420 heSerSerSerG1uAspAlaLeuG1uTyrTyrLysProSerThrThrleuAspG1yThra 440
Db 1847 TCTCTCATCAAGAAATGCTTTGGAAATCTTACAAACCTTCCACTGAGATGGAGCA 1906
Qy 440 snLysleuCysG1nPheserProValG1nGluLysSerG1uG1nThrProGluThrSerP 460
Db 1907 ACAAGCTATGCAAGTCTCCCTGTTCAGGAACATTCGAGAGCAAGCTCCGCAAAACAATC 1966
Qy 460 roAspLysG1uGluLysSer11eProLysLysleuG1nThr1aArgProSerAspSerG 480
Db 1967 CTGAATPAGGAGAGACACAGCATCCCAAGAACTCCAGACCCGCGCTTCAGACAGCC 2026
Qy 480 1nSerLysArgLeuHisSerValaArgThrSerSerSerG1yThra1aG1nThrSerleu 500
Db 2027 AGAGCAAGGATGTCATTCGTGCTCAACACAGAGAGTGGACCCGCCCAAGAGTCTCTT 2086
Qy 500 eusSerProleuHisArgSerG1ySerValG1uAspAsnTyrHisThrSerPheleuPheG 520
Db 2087 TATCTCCACTGTCATGGAAGTGGAGCGTGGAGGACAAATTCACACAGCGCTTCTCTTCG 2146
Qy 520 1yleuSerThrSerG1nHisleuThrLysSerAlaG1yLeuG1yleuLysG1yTTPH 540
Db 2147 GCTTTTCCACAGCAGCAGACCTTCAAGAGTCTGCGGCTGGGCTTAAAGGCGCTGGC 2206
Qy 540 1sSerAsp11eleuAlaProG1nThrSerThrProSerleuThrSerSerThrPlyrPhea 560
Db 2207 ACTGGATATCTGGCCCCCAGACCTTACCTTCCCTTAACAGACAGCTGTGATTTTG 2266

Qy 560 1aThrGlusSerSerHisPheTyrSerAlaSerAla11eTyrG1yG1ySerAlaSerTyrS 580
Db 2267 CCACAGAGTCTCCACACTTACTCTGCTCGCATCTACAGAGGAGAGTGCAGATTA 2326
Qy 580 eRAlaTyrSerCysSerG1nLeuProThrCysG1yAspG1nValTyrSerValaArgAga 600
Db 2327 CTGCTTACAGCTGACAGCAGCTGCCCACTTGCGAGAACCAAGTCTATTTCTGTGGCAGGC 2386
Qy 600 1rgG1nLysProSerAspAArg1aAspSerAArgSerTyrPH1sG1uG1uSerPropheG 620
Db 2387 GGCAGAGCAAGTGAACAGAGTGACTCGCGCGAGCTGGCATGAAGAGAGCCCTTTG 2446
Qy 620 1uLysG1nPhelyaAArgAArgSerCysG1nMetGluPheG1yG1uSer11eMetSerG1uA 640
Db 2447 AAAAGCACTTTAAAGCAGAAAGCTGCGCAAAATGGAATTTGAGAGAGCATATGTCAGAGA 2506
Qy 640 snAArgSerAArgG1uG1uLeuG1yLysValG1ySerG1nSerSerPheSerG1ySerMetG 660
Db 2507 ACAAGTCAAGGAGAGAGCTGGGGAAGTGGCAAGTCACTTACGCTTTTCGGGCAAGATGG 2566
Qy 660 1u11e11eG1uValSer 665
Db 2567 AAATCATTGAGAGTCTCC 2583
RESULT 23
ABV26923
ID ABV26923 standard; cDNA; 5145 BP.
XX
XX ABV26923;
XX
DT 16-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 26914.
XX
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
XX
XX Human; pharmacologic marker; gene; ss.
XX
OS Homo sapiens.
XX
XX WO200160860-A2.
XX
XX
PD 23-AUG-2001.
XX
XX
PE 20-FEB-2001; 2001WO-US005171.
XX
XX
PR 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PI Schlegel R, Endege WO, Monahan JE;
XX
XX WPI; 2001-66795/76.
XX
XX
XX Novel isolated nucleic acid molecule associated with cancerous state of
XX prostate cells and correlating with presence of prostate cancer, useful
XX for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS Claim 1; Page 5449-5450; 11750pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule (1) comprising
XX a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
XX specification or its complement. (1) is useful for: (a) assessing whether
XX a patient is afflicted with prostate cancer; (b) monitoring the
XX progression of prostate cancer in a patient; (c) assessing the efficacy
XX of a test compound to inhibit prostate cancer in a patient; (d) assessing
XX the efficacy of a therapy for inhibiting prostate cancer in a patient;
XX (e) selecting a composition for inhibiting prostate cancer in a patient;
CC

CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or incidence of prostate cancer in a patient
CC ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
SQ Sequence 5145 BP, 1408 A; 1135 C; 1253 G; 1346 T; 0 U; 3 Other;

Alignment Scores:

Pred. No.:	0	Length:	5145
Score:	472.00	Matches:	663
Percent Similarity:	99.40%	Conservative:	0
Best Local Similarity:	99.40%	Mismatches:	2
Query Match:	70.98%	Indels:	4
DB:	5	Gaps:	0

US-10-029-345A-109 (1-665) x ABV26923 (1-5145)

QY 1 MetAlaHIEGJUmEtIEGJYThrGlnIIEValThrGluArgPheValAlaLeuLeuGlu 20
DB 589 ATGGCCCATGAGATGATGGAACTCAAAATTGTTACTGAGAGGTGGTGGCTGCTGGAA 648
QY 21 SerGIYThrGluLYsValLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 40
DB 649 AGTGAACCGAAGAAAGTGGCTGTAATTGATAGCCGGCCATTGGAAATACATATACCTC 708
QY 41 HIEIIEuGluAlaIIEaenIIEaenCYSerLYsLeuMetLYsArgIIEuGlnGln 60
DB 709 CACATTTTGGAAACCATTAATATCAACTGCTCCAACTTATGAAAGGAAGTGGCAACG 768
QY 61 AspLYsValLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 80
DB 769 GACAAAGTGTATTCAGAGCTCATCCAGATTCCAGCAACATAGGTTGACATTGAT 828
QY 81 CYSerGlnLYsValValValValValValValValValValValValValValValVal 100
DB 829 TGCAGTCAGAGGTGATGATTCAGATCAAAAGCTCCCAAGATGTTGCTCTCTCTCA 888
QY 101 AspCYsPheLeuThrValLeuLeuGlyLYsLeuGlnLYsSerPheLeuLeuValHIEu 120
DB 889 GACTGTTTCTCACTGACTTCTGGGTAACTGGAGAAAGCTTCAACTCTGTTCACTG 948
QY 121 LeuAlaGlyLYsPheAlaGluPheSerArgCYsPheProGlyLeuCYsGlnGlyLYsSer 140
DB 949 CTTCGACGGTGGTTCGAGTCTCTCTGTTTCTGCTGGCTCTGTTGAGAAATATC 1008
QY 141 ThrLeuValProThrCYsIIESerGlnProCYsLeuProValAlaAenIIEGlyProThr 160
DB 1009 ACTCTAGTCCCTACCTGCAATTTCTCAGCCTTGCTTACTGTTGCCAAATGGGCCAAC 1068
QY 161 ArgIIEuProAenLeuThrLeuGlyCYsGlnArgPheValLeuAenLYsGluLeu 180
DB 1069 CGAATCTTCCCAATCTTATCTTGGCTGCACGAGATGCTCTCAACAAAGAGCTGAT- 1127
QY 181 -GlnGlnaenGlyIIEGlyThrValIIEuAenIIEaenIIEaenIIEaenIIEaenIIEaen 200
DB 1128 GCGACGAATGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1186
QY 200 HeIIEProGlnSerHIEPheLeuArgValProValAenAenPheCYsGlnLYsIIEaen 220
DB 1187 TTATCCCGAGTCTCATTTCTCGGTGCTGCTGATGATGATGATGATGATGATGATGAT 1246
QY 220 eubProTrpLeuAenLYsSerValIIEPheIIEGlyLYsAlaLYsAlaSerAenGlyCYs 240
DB 1247 TGCCTGGTGGTGGCAATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1306
QY 240 AlLeuValHIEsValaenIIEaenIIEaenIIEaenIIEaenIIEaenIIEaenIIEaen 260
DB 1307 TTCTAGTGCATGTTAGCTGGGATCTCCCGCTCCGACACATGCTATGCTATGCTATGCT 1366
QY 260 eLYsArgMetSerLeuAenGlyValIIEaenIIEaenIIEaenIIEaenIIEaenIIEaen 280
DB 1367 TGAAGAGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1426

QY 280 HIEIIESerProAenPheAenPheLeuGlyGlnLeuLeuAenLYsGlyLYsIIEaen 300
DB 1427 CTATATCTCCAACTTCAATTTTCTGGCCAACTCTGATCTATGAGAAAGATTTAAGA 1486
QY 300 eGlnThrGlyAlaSerGlyProLYsSerLYsLeuLYsLeuLeuLYsGlyLYsPro 320
DB 1487 ACCAGCTGAGCATCAGGCGCCAAAGAGAACTCAAGCTGCTGACCTGGAGAACCA 1546
QY 320 eGlnProValProAlaValaSerGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 340
DB 1547 ATGAACCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1606
QY 340 roCYsAlaAenPheIIEaenIIEaenIIEaenIIEaenIIEaenIIEaenIIEaenIIEaen 360
DB 1607 CCTGTGCGACTCTGCTACTGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1666
QY 360 AlProSerValProSerValGlnProSerLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 380
DB 1667 TCCCAAGGTGCGCCAGCGTGCAGCGCTGCTGTTAGAGAGACCGCTGTTACAGGCC 1726
QY 380 eusErGIYLeuHIEsSerAlaAenArgLeuGlnAenPheAenLYsLeuLYsArgSer 400
DB 1727 TCAGTGGCTGCACCTGCTCCGAGACAGGCTGGAAGACATAGCTCAAGCTTCTCT 1786
QY 400 HeSerLeuAenPHEIIEsSerValaSerLYsSerAlaSerLeuHIEaenIIEaenIIEaen 420
DB 1787 TCTCTGATATCAATCATGTTTCAATTACGACGACATGAGATCTTCAATCATAGGCT 1846
QY 420 HeSerSerSerGlnAenAlaLeuGlyLYsLYsPheSerThrThrLeuAenGlyThra 440
DB 1847 TCTCTCATCAAGAATGCTTTGGAATACAAACCTTCCACTGATGGAGACCA 1906
QY 440 eLYsLeuCYsGlnPheSerProValGlnGluLeuSerGlnGlnThrProGluThrSer 460
DB 1907 ACAAGCTATGACAGTCTCTCCCTGTCAGAACTATCGAGAGACATCCCGAAACCA 1966
QY 460 roAenLYsGlnGlnIIEaenIIEaenIIEaenIIEaenIIEaenIIEaenIIEaenIIEaen 480
DB 1967 CTGATTAAGAGAGAGCAGACATCCCAAGAGCTGACAGCCGACAGCTTTCAGAGACC 2026
QY 480 IIEsSerArgLeuHIEsSerValArgThrSerSerSerGlyThraIIEaenIIEaen 500
DB 2027 AAGAGAGAGATGATGATTCGTCAGAACCAAGACAGAGCTGACCCGACAGAGTCTCT 2086
QY 500 eusErProLeuHIEsArgSerGlySerValGlnAenAenThrHIEsSerPheLeuPhe 520
DB 2087 TATCTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2146
QY 520 IIEuLeuSerThrSerGlnGlnIIEaenIIEaenIIEaenIIEaenIIEaenIIEaenIIEaen 540
DB 2147 GCTTTTCCACAGCCAGCAGACCTTACGAAAGTCTGCTGCTGCTGCTTAAAGGCTG 2206
QY 540 IIEsSerAenIIEaenAlaProGlnThrSerThrProSerLeuThrSerSerThrPhe 560
DB 2207 ACTCGATATCTTGGCCCCCAGACCTTACCCCTTCCCTGACAGCAGACTGGTATTTG 2266
QY 560 IIEaenIIEaenIIEaenIIEaenIIEaenIIEaenIIEaenIIEaenIIEaenIIEaen 580
DB 2267 CCACAGAGCTTCACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2326
QY 580 eRLaIIEsSerCYsSerGlnLeuProThrCYsGlyAenPheValIIEsSerValArgArg 600
DB 2327 CTGCTTACAGCTGACAGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2386
QY 600 rGlnLYsProSerAenPheArgAlaAenPheArgSerThrHIEaenIIEaenIIEaenIIEaen 620
DB 2387 GCGAGAGCCAGTACAGAGCTGATCGCGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCT 2446
QY 620 IIEuLYsGlnPheLYsArgSerCYsGlnMetGlnPheGlyGlnSerIIEaenIIEaenIIEaen 640
DB 2447 AAAAGAGTTTAAAGCAGAGCTGCAATGGAATTTGGAGAGAGCATGATGAGAGA 2506
QY 640 eHIEsSerArgGlnGluLeuGlyLYsValIIEsSerIIEsSerPheSerGlySerMetG 660

Db 2507 AAGGTCACGGGAAGAAGCTCGGGGAAATGGGCAGTCAAGTTTGCGGACAGCATGG 2566
Oy 660 lulleilegiva1ser 665
Db 2567 AAATCATTGAGGTCTCC 2583

RESULT 24
ABV27135
ID ABV27135 standard; cDNA; 5145 BP.
XX AC ABV27135;
XX DT 16-SEP-2002 (first entry)
DE Human prostate expression marker CDNA 27126.
XX KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.
XX OS Homo sapiens.
XX PN WO200160860-A2.
XX PD 23-AUG-2001.
PF 20-FEB-2001; 2001WO-US005171.
PR 17-FEB-2000; 2000US-0189319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PI Schlegel R, Endege WO, Monahan JE;
DR WPI; 2001-662795/76.
XX PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer; useful
PT for detecting presence of prostate cancer; stage of prostate cancer.
PS Claim 1; Page 5502-5503; 11750pp; English.

The invention relates to an isolated nucleic acid molecule (I) comprising
a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
specification or its complement. (I) is useful for: (a) assessing whether
a patient is afflicted with prostate cancer; (b) monitoring the
progression of prostate cancer in a patient; (c) assessing the efficacy
of a test compound to inhibit prostate cancer in a patient; (d) assessing
the efficacy of a therapy for inhibiting prostate cancer in a patient;
(e) selecting a composition for inhibiting prostate cancer in a patient;
(f) assessing the prostate cell carcinogenic potential of a compound; (g)
determining whether prostate cancer has metastasized in a patient; (h)
assessing the aggressiveness or indolence of prostate cancer in a patient
; (I) is also useful as a pharmacodynamic or pharmacogenomic marker

Sequence 5145 BP; 1408 A; 1135 C; 1253 G; 1346 T; 0 U; 3 Other;

Alignment Scores:
Pred. No.: 0 Length: 5145
Score: 472.00 Matches: 663
Percent Similarity: 99.40% Conservative: 0
Best Local Similarity: 99.40% Mismatches: 2
Query Match: 70.98% Indels: 4
DB: 5 Gaps: 0

US-10-029-345A-109 (1-665) x ABV27135 (1-5145)

I MetalahisglumetlelglyThrglnlleva1ThrgluArgLeuValAlaLeuLeugln 20

Db			ATGGCCCAAGAAATGATTGGAACTCAAATGTATTACGAAGGTTGGTGCCTCGCGGAA	648
Oy	21	SerGIYThGlulysValIleuLeuIleAspSerArgProPheValJalunTYraenThrSer		40
Db	649	AGTGAAACGAAAAAGTCGTCTAATTGATTAGCCGCCCATTTGTGGAAATRCATAATCACC		708
Oy	41	HsiIeLeuGluAlaIleasnIleasncySserLYsleuMetlysaRghrIeugIngin		60
Db	709	CACATTTTGGAGCCATTATATCAACTGCTCCAAGCTTATGAAGCGAAGTTGCAACAG		768
Oy	61	AaplylValIleuIIeThrgIleuIleGINHIseseRIalyehIslysValIleApIleAsp		80
Db	769	GACAAAGGTTAATTATACAGAGCTCATCCAGCATTCAGCAAACTATAGAGTTGCAATTGAR		828
Oy	81	CysSerGlnLysValValValTYraSpGINserSerGlnAspValAlaSerLeuSerSer		100
Db	829	TCCAGTCAGAAGGTTGTATTATACATCAAGCTCCCAGAAATGTTGCCCTCTCTCTTCA		888
Oy	101	AapCySpheLeuThrValIleuLeuGlyLYsleuGluLYsSerPheAsnSerValHIslEu		120
Db	889	GACTGTTTTCTACGTACTTCTGGGTAAACTGGAGAAAGAGCTTCAACTCTGTTCACTG		948
Oy	121	IeuAlaglyGLYPheAlaglupheseraGcSphProglyLYeucYsgIuGlyLYsSer		140
Db	949	CTTGAGAGGGGTTGCTGAGATTCTCGTTGTTTCCCGGCTCGTGAAGAAAATCC		100
Oy	141	ThrlieuValProThrCYseIleSerGlnProCYsleuProValAlaSnIlegIProthr		160
Db	1009	ACTCTAGTCCCTAACCTGCATTTCTCAGCCTTCTTACCTGTTGCCAACATTTGGGCCAAC		106
Oy	161	ArgIleLeuProAsnLeuTyrldeuGlyCYsGlnArGaSPValIleuSnlySGIuleuIle		180
Db	1069	CGAATTTCTTCCCAATCTTATCTTGAGCTGCCAGCGAGATGTCCTCAACAAGAGCTGAR-		112
Oy	181	-GInglInengLyIllegLYrrValIleuAsnAlaserTYr-ThrCYsProLYsProAsp		200
Db	1128	GGAGAGAAATGGAAATGGTTAGTTAGTTAAATCCAGCA-TRCCGTGCCAAAGCCGAGCT		118
Oy	200	heIlProcluserIAsPheIeuArGVAlProValAsnaSPserPheCYsGIuLYsIlEl		220
Db	1187	TTATCCCCGAGCTCAATTTCTCGCTGCGCTGTGAATGACAGCTTTTGTGAAANAATTT		124
Oy	220	eupProTrIpLeuAspLYsSerValJaSPheIllegIuLYsAlaLYsAlaseRaEnglyCYsV		240
Db	1247	TGCCGTGGTGGCAAAATCAGTAGATTTCATTGAGAAAGCAAAAGCCTCCAAATGATGIG		130
Oy	240	alleuValHisCYsleuAlaglyIleserARgsSerAlaThrlIleaIIeaIATryIlem		260
Db	1307	TTCTGTGTGACACTGTTAGCTGGGAACTCCCGCTCCGCCCACTGCTATCGCTTCATCA		136
Oy	260	eLYsarIMetAspMetSerIleuaSPGJualATyrArPheValLYsGIuLYsARgProT		280
Db	1367	TGAAGAAGATGGACATGTCTTTAGATGAAGCTTACGATTGTGAAAGAAAAGACCTTA		142
Oy	280	hrILeserProAsnPhesnPheLeuGlyGlnLeuLeuAsPTyrGIuLYsLYsIIelySA		300
Db	1427	CTATATCTCCAAACTTCAATTTTCTGGGCAACTCTGGACATATAGAAAGAAATTAAATA		148
Oy	300	enGIInThrGlyAlaSerGIyProLYsSerLYsLeuLYsIleuLeuHIsleuGIuLYsProA		320
Db	1487	ACCAAGACTGGAGCATCACGGCCAAAGACCAAACTCAAGCTCTGCACCTGGAGAAAGCCA		154
Oy	320	enGIuProValProIlaValSerGIuGlyGIuGlnLYsSerGluNhrProIleuSerProp		340
Db	1547	ATGAACCTGTCCCTGCTGCTCAGAGGGGAGCACAAAAAAGCAAGCCCTCAAGTCCAC		160
Oy	340	roCYsAlaASPserAlaThrSerGIuAlaAlaglyGlnArGPProValHisProAlaServ		360
Db	1607	CCTGTGCCGACTCTCTACCTCAAGGCGACAGACAAAGCCCTGTGATCCCGCCAGGG		166
Oy	360	aIProSerValProSerValGlnProSerLeuLeuGluAspSerProLeuValGlnAlaL		380

```

Db      1667 TGCCACGCTGCGCCAGCTGACCGCTGCTTTAGAGGACAGCCGCTGTACAGAGCCG 1726
Qy      380 eusecrlyleuHlslseuSerAlaAspArgleuGluAspSerAnlysluylValArgSerP 400
Db      1727 TCAGTGGGCTGCACCTGTCCGACAGACGCTGGAAACAGCAATAGCTTCAGCGTTCTCT 1786
Qy      400 hseSerleuAspIlelySerValSerTyrSerAlaSerleuAlaAspSerleuHlslgIyp 420
Db      1787 TCTCTTGATATCAATCAAGTTCTCATATTACAGCCAGCATGGACATCTTACATAGGCT 1846
Qy      420 hseSerSerGluAspAlaleuGluTyrTyrlyAspProserThrThrleuAspGlyThra 440
Db      1847 TCTCCATCAGAAAGATGCTTTGGAAATCATCAAACTTCCTCACTAGATGGAGCA 1906
Qy      440 enlyleuCyseGlnpHeSerProvalGlnGluSerGluGlnThrProGluThrSerP 460
Db      1907 ACAAGCTATGCCAGTTCCTCCCTGTACAGAACTATCGGAGCGAGATCCCGAAACAGTC 1966
Qy      460 roAspIlyseGluAlaSerIleProlylyleuGlnThrAlaArgProserAspSerG 480
Db      1967 CTGATTAAGAGAGAGCCAGCATCCCAAGAGCTGGACCGCCAGAGCTTCAGACAGCC 2026
Qy      480 lnserysArgleuHlslserValArgThrSerSerSerGlyThrAlaGlnArgSerleu 500
Db      2027 AAGAGAAAGGATTGATTCGCTCAGAAACAGACAGCATGGCACCGCCAGAGGCTCCCTTT 2086
Qy      500 euserProleuHlslargSerGlySerValGluAspAnlyslslThrSerPheleuPheG 520
Db      2087 TATCTCCACTGCATCCAGAGGGAGCGTGGAGAGCAATTAACACACAGCTTCCTTTTGC 2146
Qy      520 lylseuSerThrserGlnGlnHlslleuThrlySerAlaGlyleuGlyleuValylTrpH 540
Db      2147 GCTTTTCACACACACAGACACCTCACAGAGTCTGCTGGCTTAAAGGCTGGC 2206
Qy      540 llserysPleleuAlaProGlnThrSerThrProserleuThrSerSerTyrPheA 560
Db      2207 ACTCGATATCTTGGCCCCCAGACCTTAACCTTCCTCGACACAGAGCTGTATTTTG 2266
Qy      560 lathrlGluSerSerHlslpHeTyrSerAlaSerAlaIleTyrGlyGlySerAlaSerTyrS 580
Db      2267 CCACAGAGTCTCACACTTACTCTGCTGCTCACCCATTAACGAGGACAGTGCACGTTACT 2326
Qy      580 eeralatYrSerCyseSerGlnleuProThrCyseGlyAspGlnValTyrSerValArgArg 600
Db      2327 CTGCTTACAGCTGCACAGCTGCTCCACTTGCAGAAACAGATCTATCTGTCGCAAGC 2386
Qy      600 rGlnlyAspProserAspArgAlaAspSerArgArgSerTyrPheGlnGluSerProPheG 620
Db      2387 GCGAGAAAGCCAAAGTGAACAGCTGATCCGCGGCGAGCTGGCATGAAGAGAGCCCTTTG 2446
Qy      620 lulyeGlnpHeLyAspArgSerCyseGlnMetGluPheGlyGluSerIleMetSerGluA 640
Db      2447 AAAAGAGATTAAACGACAGAGCTGCCAAATGGAAATTTGGAGAGACATCATGTGAGA 2506
Qy      640 sAspArgSerArgGluGluGluGlylyValGlySerGlnSerSerPheSerGlySerMetG 660
Db      2507 ACAGGTACAGGAGAGAGCTGGGAGAAAGTGGAGTCACTTACCTTTTGGGACAGATGG 2566
Qy      660 lullelleGluValSer 665
Db      2567 AAATCATTAAGTCTCC 2583

```

```

KW      pharmacogenomic marker; gene; ss.
XX
OS      Homo sapiens.
XX
PN      WO200160860-A2.
XX
PD      23-AUG-2001.
XX
PF      20-FEB-2001; 2001WO-US005171.
XX
PR      17-FEB-2000; 2000US-0183119P.
PR      16-MAR-2000; 2000US-0189862P.
PR      25-MAY-2000; 2000US-0207454P.
PR      09-JUN-2000; 2000US-0211314P.
PR      18-JUL-2000; 2000US-0219007P.
PR      13-DEC-2000; 2000US-0255281P.
XX
PA      (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PI      Schlegel R, Endege WO, Monahan JB;
XX
DR      WPI, 2001-662795/76.
XX
PT      Novel isolated nucleic acid molecule associated with cancerous strate of prostate cells and correlating with presence of prostate cancer, useful for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS      Claim 1; Page 6009-6010; 11750pp; English.
XX
CC      The invention relates to an isolated nucleic acid molecule (I) comprising a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the specification or its complement. (I) is useful for: (a) assessing whether a patient is afflicted with prostate cancer; (b) monitoring the efficacy of a test compound to inhibit prostate cancer in a patient; (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient; (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient; (e) selecting a composition for inhibiting prostate cancer in a patient; (f) assessing the prostate cell carcinogenic potential of a compound; (g) determining whether prostate cancer has metastasized in a patient; (h) assessing the aggressiveness or indolence of prostate cancer in a patient; (i) is also useful as a pharmacodynamic or pharmacogenomic marker
XX
SQ      Sequence 5145 BP; 1408 A; 1135 C; 1253 G; 1346 T; 0 U; 3 Other;
XX
XX
Alignment Scores:
Pred. No.: 0
Score: 472.00
Percent Similarity: 99.40%
Best Local Similarity: 99.40%
Query Match: 70.98%
DB: 5
Gaps: 0
US-10-029-345A-109 (1-665) x ABV28657 (1-5145)
Qy      1 MetAlaHlslGluMetIlelyThrGlnIleValThrGluArgleuValAlaleuLeuGlu 20
Db      589 ATGGCCCAAGATGATGTAAGTCAAAATTTGACTGAGAGGTGTGGCTCTGCTGGAA 648
Qy      21 SerGlyThrGluValleuLeuIleAspSerArgPProPheValGluTyrAspThrSer 40
Db      649 AGTGAACGGAAAAAGCTGCTGCTAATGTATAGCCGGCCATTTGTGAAATCAATATCATCC 708
Qy      41 HlslleuGluAlaIleAsnIleAenCyseTyrlyleuMetIlyAspArgleuGlnGln 60
Db      709 CACATTTGGAAGCCATTAATATCACTCTCCAAAGCTTATAGAGCGAAGTTGCAACAG 768
Qy      61 AspIlyValleuIleThrGluIleuIleGlnHlslserAlaIlysllyValAspIleAsp 80
Db      769 GACAAAGTGTATTAATACAGAGCTCATCCAGATTCAGCAAAACATDAAGTGAATGAT 828
Qy      81 CyseSerGlnlyValValValTyrAspGlnSerSerGlnAspValAlaSerleuSerSer 100
Db      829 TGCAGTCAGAGAGTGTATGTTTATTCATCAATCAAAAGCTCCCAAGAGTTCCTCTCTCTTCA 888

```

RESULT 25
 ABV28657
 ID ABV28657 standard; cDNA; 5145 BP.
 XX
 AC ABV28657;
 XX
 DT 16-SEP-2002 (first entry)
 XX
 DE Human prostate expression marker cDNA 28648.
 XX
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;

QY 101 AspCysPheLeuThrValLeuLeuGlyValLeuGlyLysSerPheAsnSerValHisLeu 120
 DB 889 GACGTTTCTCACTGACTTCTGGGTAACGTGAGAGAGCTTCAACTCTGTTCACCTG 948
 QY 121 LeuValGlyValPheValGluPheSerArgCysPheProGlyLeuGlyGlySer 140
 DB 949 CTTCAGAGTGGGTTGCTGAGTTCTCTGTTGTTCCCTGGCTCTGTGAAAGAAATCC 1008
 QY 141 ThrLeuValProThrCysHisSerGlnProCysLeuProValAlaAsnHisLeuGlyProThr 160
 DB 1009 ACTCTAGTCCCTACCTGCATTTCTCAGCCTTGCTTACCTGTTCACCAATGGGCCAAC 1068
 QY 161 ArgHisLeuProAsnLeuThrLeuGlyCysGlnArgAspValLeuAsnLysGluLeu 180
 DB 1069 CGAATCTTCCCAATCTTTATCTTGCTGCCAGCAGAGTCTCTCAACAGAGCTGAT- 1127
 QY 181 -GlnGlnAsnGlyValGlyValValLeuAsnHisSerVal-ThrCysProLysProAsp 200
 DB 1128 GCAGCAGAAATGGATGGTATGTATTAATGCCAGCAA-TACCTGTCCAAAGCTGACT 1186
 QY 200 HeHisProGluSerHisPheLeuArgValProValAsnAspSerPheCysGluValLeu 220
 DB 1187 TTATCCCGAGTCTCATTTCTCGTGTGCTGTGAATGACAGCTTTGTGAGAAATTT 1246
 QY 220 eubProTrpLeuAspLysSerValAspPheHisGlyLysAlaLysAlaSerAsnGlyCysVal 240
 DB 1247 TGCCGTTGGTGGCAAAATCGTAGATTTTCATTGAGAAAGCAAAAGCTCAATGATATG 1306
 QY 240 AlLeuValHisCysLeuAlaGlyValSerArgSerAlaThrHisLeuAlaValLeu 260
 DB 1307 TTCTATGTGACGTGTATGCTGGGATCTCCGCTCCGCCACCATCGCTATCGCTACATCA 1366
 QY 260 eCysArgMetAspMetSerLeuAspGluAlaThrArgPheValLysGluLysArgPro 280
 DB 1367 TGAGAGGATGAGCATGCTCTTATGATGAAGCTTACGATTTGTAAGAAAGAAAGCCCTA 1426
 QY 280 hHisSerProAsnPheAsnLeuGlyValLeuLeuAspTrpGluLysHisLeuVal 300
 DB 1427 CTATATCTCCAAACTTCANATTTCTGGGCCAACTCTGAGCACTATGAGAAAGATTAAGA 1486
 QY 300 sngGlnThrGlyAlaSerGlyProLysSerLysLeuLysLeuHisLeuGlyLysPro 320
 DB 1487 ACCAGACTGAGCATCAGGCGCAAGAGCAAACTCAAGCTGTGACCTGGAGAGCCAA 1546
 QY 320 sngGluProValProAlaValSerGluGlyValGlyLysSerGlnThrProLeuSerPro 340
 DB 1547 ATGAACCTGTCCCTGTGTCTCAGAGGTGAGACAGAAAGCGAGCGCCCTCAAGTCCAC 1606
 QY 340 roCysAlaAspSerAlaThrSerGluAlaAlaGlyValArgProValHisProAlaSer 360
 DB 1607 CCGTGTGCCAGCTCTGCTACCTCAGAGGCGAGGCAAAAGCCCGTGCATCCGCGACAG 1666
 QY 360 AlProSerValProSerValGlnProSerLeuLeuGluAspSerProLeuValGlnAla 380
 DB 1667 TCCCGAGCTGTGCCAGCGCGTCTGTGTGAGAGACGCCCGCTGTGACAGCGCC 1726
 QY 380 eusSerGlyLeuHisLeuSerAlaAspArgLeuGluAspSerAsnLysLeuLysArgSer 400
 DB 1727 TCAGTGGGTGACCTGTCCGAGACAGGCTGAAAGACGCAATAGCTCAAGCGTTCTCT 1786
 QY 400 hSerLeuAspHisLysSerValSerValSerAlaSerMetAlaHisSerLeuHisGly 420
 DB 1787 TCTCTCTGATATCAAAATCATATTCAATTCAGCCAGCATGGCAGCATCTTACATGCT 1846
 QY 420 hSerSerSerGluAspAlaLeuGlyValLysProSerThrThrLeuAspGlyThr 440
 DB 1847 TCTCTCTCATGAAAGATCTTTGGAATCTCAAACTTCACTCTGTGATGGAGCA 1906
 QY 440 snLysLeuGlyGlnPheSerProValGlnGluLeuSerGlyGluThrProGluThrSer 460
 DB 1907 ACAAGCTATGCGAGTTCTCCCTGTTCAGGAATCTCGAGAGCACTCCGAAACAGATC 1966

QY 460 roAspLysGluGluAlaSerHisProLysLysLeuGlnThrAlaArgProSerArgSer 480
 DB 1967 CTGATTAAGAGGAACCGACATCCCAAGAGCTGACAGCGCGGCTTTCAGACAGC 2026
 QY 480 InSerLysArgLeuHisSerValArgThrSerSerSerGlyThrAlaGlnArgSerLeu 500
 DB 2027 AGAGCAGAGATTCATTCGGTCAGAACAGAGAGAGTGGACCGCCAGAGTCCCTTT 2086
 QY 500 eusSerProLeuHisArgSerGlySerValGluAspAsnTrpHisThrSerPheLeuPhe 520
 DB 2087 TATCTCCATGATCGAAGTGGAGCTGTGAGAGCAATTAACACACAGCTTCTTTTCG 2146
 QY 520 LysSerThrSerGlnGlnHisLeuThrLysSerAlaGlyLeuGlyLeuLysGlyTrp 540
 DB 2147 GCCTTTCACAGCCAGCAGACCTTCAGAAAGTCTGTGCTGGCTTAAAGGCTGCG 2206
 QY 540 IsSerAspHisLeuAlaProGlnThrSerThrProSerLeuThrSerSerTrpThrPhe 560
 DB 2207 ACTGGATATCTTGCCCCCGAGACTTACCTTCCCTGACAGCAGCTGTATTTTG 2266
 QY 560 LathGluSerSerHisPheTrpSerAlaSerAlaLeuTrpGlyLysSerAlaSerTrp 580
 DB 2267 CCACAGAGTCTTCACACTTCTACTCTGCTCAGCCATCTACGAGGCGAGTCCAGTTACT 2326
 QY 580 eAlaTrpSerCysSerGlnLeuProThrCysGlyAspGlnValTrpSerValArgArg 600
 DB 2327 CTGCTTACAGCTGACCCAGCTGCCACTTCCGAGACCAAGTCTATCTGTGTGGCAGGC 2386
 QY 600 TrgGlnLysProSerAspArgAlaAspSerArgArgSerTrpHisGlyLysSerProPhe 620
 DB 2387 GGCAGAGCCCAAGTACAGAGACTGACTCGCGCGAGCTGGCATGAAAGAGCCCTTTG 2446
 QY 620 LysGlnPheLysArgArgSerCysGlnMetGluPheGlyLysSerHisLeuSerGlu 640
 DB 2447 AAAAGCACTTAAACGAGAGCTGCAAAATGAAATTTGAGAGAGCATCATGTCAAGA 2506
 QY 640 snArgSerArgGluGlyLeuGlyLysValGlySerGlnSerSerPheSerGlySerMet 660
 DB 2507 ACAGTCACTGGAGAGACTGGGGAAGTGGGCACTAGTCTAGCTTTTGGGCGCATGG 2566
 QY 660 LulleHisGluValSer 665
 DB 2567 AAATCATTTGAGGTCTCC 2583
 ID ABV22827 standard; cDNA: 5145 BP.
 AC ABV22827;
 XX 13-SEP-2002 (first entry)
 DT Human prostate expression marker cDNA 22818.
 DE Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
 KW pharmacogenomic marker; gene; se.
 XX Homo sapiens.
 OS Homo sapiens.
 XX MO200160860-A2.
 PN MO200160860-A2.
 PD 23-AUG-2001.
 PF 20-FEB-2001; 2001MO-US005171.
 XX 17-FEB-2000; 2000US-0183319P.
 PR 16-MAR-2000; 2000US-0189862P.
 PR 25-MAY-2000; 2000US-0207454P.
 PR 09-JUN-2000; 2000US-0211314P.
 PR 18-JUL-2000; 2000US-0219007P.
 PR 13-DEC-2000; 2000US-0255281P.
 XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Schlegel R, Endege WO, Monahan JB.
XX WPI, 2001-662795/76.
XX Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX Claim 1, Page 4033-4034, 11750pp, English.
XX
XX The invention relates to an isolated nucleic acid molecule (1) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (1) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (1) is also useful as a pharmacodynamic or pharmacogenomic marker
SQ
SQ Sequence 5145 BP; 1408 A; 1135 C; 1253 G; 1346 T; 0 U; 3 Other;
Alignment Scores:
Pred. No.: 0 Length: 5145
Score: 472.00 Matches: 663
Percent Similarity: 99.40% Conservative: 0
Best Local Similarity: 99.40% Mismatches: 2
Query Match: 70.98% Indels: 4
DB: 5 Gaps: 0
US-10-029-345A-109 (1-665) X ABV22827 (1-5145)
QY 1 MetAlaHisGluMetIleGlyThrGlnIleValThrGluArgLeuValAlaLeuLeuGlu 20
DB 589 ATGGCCCAATGAGTATGAGTGAAGTCAAAATTTGTTACTGAGAGTGGTGGCTGCTGAGAA 648
QY 21 SerGlyThrGluLeuValLeuLeuIleAspSerArgProPheValGluTyrAsnThrSer 40
DB 649 AGTGAACCGGAAAGTGGCTGCTAATGATAGCCGCGCATTTTGTGAATACATAATCATC 708
QY 41 HisIleLeuGluAlaIleAsnIleAsnCySerLeuLeuMetIleArgArgLeuGlnGln 60
DB 709 CACATTTTGGAAAGCCATTAATATCAACTGCTCCAACTTAATGAAAGGAGTTGCAACAG 768
QY 61 AspIleValIleuIleThrGluLeuIleGlnHisSerAlaIleHisIleValAspIleAsp 80
DB 769 GACAAAGTGTATTAACAGAGCTCATCCAGCATTCAGCGAACAATAGGTTGACATTTGAT 828
QY 81 CySerGlnIleValValValIleTyrAspGlnSerSerGlnAspValAlaSerLeuSerSer 100
DB 829 TCGAGTCAGAGGTTGTTAGTTTACATCAAGCTCCCAAGATGTTCCCTCTCTCTTCA 888
QY 101 AspCyPheLeuThrValLeuLeuGlyIleLeuGlnIleSerPheAsnSerValHisLeu 120
DB 889 GACTGTTTCTCACTACTCTCTGAGTAACTGAGAAAGAGCTTCAACTCTGTTCACTG 948
QY 121 LeuAlaGlyIlePheAlaGluPheSerArgCyPhePheProGlyLeuGlyGluGlySer 140
DB 949 CTTCGAGGGGGTTCCTGAGTCTCTCGTTGTTCCCTGAGCTCTGTGAAGAAATACC 1008
QY 141 ThrLeuValProThrCysIleSerGlnProCyLeuProValAlaAsnIleGlyProThr 160
DB 1009 ACTCTAGTCCCTACCTGCAATTTCTCAGCTTGCTTACCTGTTCCCAACATTGGGCCAAC 1068
QY 161 ArgIleLeuProAsnLeuTyrIleGlyCyGlnArgAspValLeuAsnIleGluLeu 180
DB 1069 CGAATCTTCCCAATCTTATATCTTGCTGCGCAGCGAGATGCTCTCAACAGAGGCTGAT - 1127
QY 181 -GlnGlnAsnGlyIleGlyTyrValLeuAsnIleSerTyr-ThrCySProLeuProAsp 200

DB 1128 GCAGCAGATGGGATTGGTTATGTGTATATCCACGAA-TACCTCTCCAAAGCCGACT 1186
QY 200 heIleProGlnSerHisPheLeuArgValProValAsnAspSerPheCyGluIleLeu 220
DB 1187 TTATCCCGAGCTCATTTCTCGCGTGCCTGTGAATACAGACTTTTGTGAAATATT 1246
QY 220 euProTrpLeuAspIleSerValAspPheIleGluValAlaIleAsnArgIleCyG 240
DB 1247 TCCCGGTGTGACAAATCAGTAGATTTCATTATGAGAAACCAAAAGCTCAATGATGTG 1306
QY 240 alleuValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIle 260
DB 1307 TTCTAGTGCACCTGTTAGTTAGTTGGAGTCTCCGCTCCGCAACCTGTTATCGCTACATCA 1366
QY 260 eLLysArgMetAspMetSerLeuAspGluAlaIleTyrArgPheValIleGlyIleArgPro 280
DB 1367 TGAAGAGATGACATGCTCTTATGATGAAGCTTACAGATTGTGAAAGAAAGAACCTA 1426
QY 280 hTrIleSerProAsnPheAsnPheLeuGlyGlnLeuLeuAspTyrGlyIleValIleVal 300
DB 1427 CTATATCTTCCAAACTTCAATTTTCTGGGCCAACTCTGAGCTATGAGAAAGATTMAA 1486
QY 300 nGlnThrGlyAlaSerGlyProIleSerIleLeuIleAsnIleLeuGluIlePro 320
DB 1487 ACCAGACTGACATCAAGGCCAAGAGCAAACTCAAGCTGCTGCACTGGAGAACCA 1546
QY 320 nGlnProValProAlaValSerGluGlyIleGlnIleSerGluThrProLeuSerPro 340
DB 1547 ATGAACCTGCTCCCTGCTGCTCAGAGGGTGAAGAAAGCAGAGAGCCCTTCAAGTCCAC 1606
QY 340 roCyAlaAspSerAlaThrSerGluAlaIleGlyGlnArgProValHisProIleSer 360
DB 1607 CCTGTGCCACTCTGCTACCTCAGAGCGACGAGCAAAAGCCCGTGAATCCGCGACAGG 1666
QY 360 aIProSerValProSerValGlnProSerLeuLeuGluIleAspSerProLeuValGlnAla 380
DB 1667 TGCACAGCTGCCACCGCGAGCGCGCTGTTAGAGACAGCCGCTGGTACAGGCCG 1726
QY 380 euSerGlyLeuHisIleSerAlaAspArgLeuGluIleAspSerAsnIleLeuValArgSer 400
DB 1727 TCAGTGGCTGACCTGCTCGCAGAGAGGCTGGAAGACAGCATATAGCTCAAGCTTCT 1786
QY 400 heSerLeuAspIleIleSerValSerTyrSerAlaSerMetAlaIleSerLeuHisGly 420
DB 1787 TCTCTGTGATATCAAAATAGTTTCAATATTCAGCCAGCAATGACAGATCTTACATGCT 1846
QY 420 heSerSerSerGlnAspAlaLeuGluTyrTyrIleProSerThrThrLeuAspGlyThr 440
DB 1847 TCTCTCATCAAGAGTGTCTTGAGATCTACAAACCTTCCACTACTCTGGATGGACCA 1906
QY 440 nIleLeuCyGlnPheSerProValGlnIleLeuSerGluGlnThrProGluThrSer 460
DB 1907 ACAAGTATGACAGTTCTCCCTGTTACAGAACTATCGAGAGACTCCGAAACCAAGTC 1966
QY 460 roAspIleGluGluAlaSerIleProIleGlyLeuGlnThrAlaArgProSerAspSer 480
DB 1967 CTGATTAAGAGAAAGCCACATCCCAAGAACTGACAGCCGCAAGGCTTACAGAGCC 2026
QY 480 InsSerIleArgLeuHisSerValArgThrSerSerSerGlyThrAlaGlnArgSerLeu 500
DB 2027 AAGCAAGGATTCGATTCGCTCAGAACCAAGACAGAGTGGACCCGCGAGAGTCCCTTT 2086
QY 500 euSerProLeuHisArgSerGlySerValGluAspAsnTyrHisThrSerPheLeuPhe 520
DB 2087 TATCTCCATGTCATGAGAGTGGAGCGTGAAGACATTAACCAACCAACTCTCTTTTGG 2146
QY 520 ILeuSerThrSerGlnGlnHisIleuThrIleSerAlaGlyIleuGlyIleuIleGlyTyr 540
DB 2147 GCTTTTCCACCGCAGCAGCACTCAGCAAGTCTGCGCTTGAAGGCTGAGC 2206
QY 540 IeSerAspIleLeuAlaProGlnThrSerThrProSerIleuThrSerSerTyrIlePhe 560

Db 2207 ACTCGATATCTTGCCCCCAGACCTCTACCCCTCTCTGACCAAGCAGCTGATTTTG 2266
Qy 560 lATHrGusSerSerHisPheTySerAlaSerAlaIleTyrglyGlySerAlaSerTyrS 580
Db 2267 CCACAGATCTCTACACTTCTACTCTGCTCAGCATCTACGAGGAGGAGCCAGTTACT 2326
Qy 580 eAlaTySerCySerGlnLeuProThrCySglAaSerGlnValTySerValArgArg 600
Db 2327 CTGCTACAGCTGACGACGCTGCCCCACTTGGGAGACCAAGCTATCTCTGCGCAGGC 2386
Qy 600 rGlnTySProSerAspArgAlaAspSerAlaArgSerTyrPHsglyGlySerProPheG 620
Db 2387 GGCAGAGCCAGAGGACAGAGCTGACTCGCGCGGAGCTGGCATGAGAGGCCCTTGG 2446
Qy 620 lutySglnPheTySArgSerCySglMetGlnPheGlyGlySerIleMetSerGln 640
Db 2447 AAAAGCATTTAAACGAGAACTGCCAAATGGAAATTTGAGAGAGCATCATCTAGAGA 2506
Qy 640 snArSerArgGlnGluLeuGlyLySValGlySerGlnSerSerPheSerGlySerMetG 660
Db 2507 ACAGGTCAACGGAGAGAGCTGGGAAAGTGGGAGCATGCTACTTTTGGGAGAGCATG 2566
Qy 660 lullelleglValSer 665
Db 2567 AAATCATGTGAGGTCTCC 2583

RESULT 27

ABV26934
ID ABV26934 standard; cDNA; 5145 BP.
XX

ABV26934;
XX

16-SEP-2002 (first entry)
XX

Human prostate expression marker cDNA 26925.
DE

Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;
KW pharmacogenomic marker; gene; ss.

Homo sapiens.
OS

MO200160860-A2.
PN

23-AUG-2001.
PD

20-FEB-2001; 2001WO-US005171.
PF

17-FEB-2000; 2000US-0183319P.
PR

16-MAR-2000; 2000US-0189862P.
PR

25-MAY-2000; 2000US-0207454P.
PR

09-JUN-2000; 2000US-021314P.
PR

18-JUL-2000; 2000US-0219007P.
PR

13-DEC-2000; 2000US-0255281P.
PR

(MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
PA

Schlegel R, Endege WO, Monahan JE;
PI

WPI; 2001-662795/76.
XX

Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
for detecting presence of prostate cancer, stage of prostate cancer.

Claim 1; Page 5453-5454; 11750pp; English.
XX

The invention relates to an isolated nucleic acid molecule (I) comprising
a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
specification or its complement. (I) is useful for: (a) assessing whether
a patient is afflicted with prostate cancer; (b) monitoring the
progression of prostate cancer in a patient; (c) assessing the efficacy
of a test compound to inhibit prostate cancer in a patient; (d) assessing
the efficacy of a therapy for inhibiting prostate cancer in a patient;

CC (e) selecting a composition for inhibiting prostate cancer in a patient;
CC (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC determining whether prostate cancer has metastasized in a patient; (h)
CC assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (i) is also useful as a pharmacodynamic or pharmacogenomic marker
XX

Sequence 5145 BP; 1408 A; 1135 C; 1253 G; 1346 T; 0 U; 3 Other;

Alignment Scores:

	Pred. No.:	Length:	Matches:
Score:	0	5145	5145
Percent Similarity:	472.00		663
Best Local Similarity:	99.40%		0
Query Match:	70.98%	Mismatches:	2
DB:	5	Indels:	4
		Gaps:	0

US-10-029-345A-109 (1-665) x ABV26934 (1-5145)

Qy 1 MetAlaHsglMetIlglyThrglnlLeValThrgluArgLeuValAlaLeuGln 20
Db 589 ATGGCCCATGAGATGATGGAACCTCAAAATGTACTGAGAGGTGGCTCTCTCGAA 648
Qy 21 SerGlyThrgluTySValLeuLeuIleAspSerArgProPheValGluTyAsnThrSer 40
Db 649 AGTGGAGCGAAGAAAAGTCTCTAATGATAGCCGCGCATTTGTGAAATACATACATCC 708
Qy 41 HisIleLeuGlnAlaIleAsnIleAsnCySerTySLeuMetTySArgLeuGln 60
Db 709 CACATTTTGAAGCATTAATATATCACTGCCAAGCTTATGAGAGCCAGGTTCCACAG 768
Qy 61 AspTySValLeuIleThrglnLeuIleGlnHisSerAlaTySValAspIleAsp 80
Db 769 GACAAAGTTAATTAACAGACTCATCCAGCATTCAGGAAACATTAAGTTGACATTTGAT 828
Qy 81 CySerGlnTySValAlaValTySAspGlnSerSerGlnAspValAlaSerLeuSerSer 100
Db 829 TGCACTCAGAGGTTGATGATTAAGATCAAAAGTCCCAAGATGTGCTCTCTTTCA 888
Qy 101 AspCySPhelLeuThrglnLeuLeuGlyLySLeuGlnTySLeuPheAsnSerValHisLeu 120
Db 889 GACTGTTTCTCAGCTGATCTTCTGGTAACTGGAGAAAGCTTCAACTGTTCACCTG 948
Qy 121 LeuAlaGlyGlyPheAlaGluPheSerArgCySPhProGlyLeuCySglTySLeuSer 140
Db 949 CTTCAGAGTGGGTGTGCTGAGTTCTCTGTTGTTTCCCTGCTCTGAGAAAGAAATCC 1008
Qy 141 ThrLeuValProThrCySLeuGlnProCySLeuProValAlaAsnIleGlyProThr 160
Db 1009 ACTCTAGTCCCTACCTGATTTCTCAGCTTGTCTTACTGTGCAACATTTGGGCCAAC 1068
Qy 161 ArgIleLeuProAsnLeuTySLeuGlyCySglAspValLeuLeuTySglTySLeu 180
Db 1069 CGAATCTTCCCAATCTTAATCTTGCTGCGCAGAGAGATCTCTCAACAGAGCTGAT- 1127
Qy 181 -GlnGlnAsnGlyIleGlyTySValLeuAsnAlaSerTyS-ThrCySProTySAsp 200
Db 1128 GCAGCAAGATGGAGTGTGATGATGATTAATGCAAGAA-TACTGTCCAAACCTGACT 1186
Qy 200 helLeProGlySerHisPheLeuArgValProValAsnAspSerPheCySglTySLeu 220
Db 1187 TTATCCCGAGTCTCAATTTCTCTGCTGCTGATGATGACACATTTTGGAGAAATTT 1246
Qy 220 eUpProTySLeuAspTySValAspPheIleGlnTySAlaTySAlaSerAsnGlyCyS 240
Db 1247 TGCCTGCTGAGCAAAATCAGATTTCACTTGAAGAAAGCAAGCTCCAAAGATGG 1306
Qy 240 alleuValHisCySLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTySLeu 260
Db 1307 TTCTAGAGCATCTTTAGCTGGGATCTCCGCTCCGCCACACATGCTATGCTTACATCA 1366
Qy 260 eTySArgTySAspMetSerLeuAspGlnAlaTySArgPheValTySglTySArgPro 280
Db 1367 TGAAGAGATGAGCATGTCTTTAGATGAAGCTTACGATTTGAGAAAGAAAGAACCTTA 1426

QY	280	nrllsErPrOaBnPhenbPheLengIyGlnLeuLeuAspTyGlnIulYulYsllElybA	300
Db	1427	CTATATCTCCAAACTTCGAATTTCTGGGCGCAACTCTTGAGACTATGAGAAAGATTAA	1486
QY	300	angInThGjVAlaSerGlyProLYsSerLYsLeuLYsLeuLeuHnIstEugIulYsProA	320
Db	1487	ACCAAGACTGGAGCATCAGGCGCCAAAGAACAACTCACTGCTGCACTGGAGAAAGCCA	1546
QY	320	angIuProVaJProAlaValSerGluGlyGlnLYsSerGluThrProLeuSerProp	340
Db	1547	ATGAACCTGCTCCTGCTGCTCTCAGAGGGGTGGACAGAAAAAGAGAGCGCCCTCAGTCCAC	1606
QY	340	roCYsAlaAspSerLYsIathSerGluAlaIagIuGlnaYsProVAlHAspAlaSerV	360
Db	1607	CGTGAGCGCACTCTCTACTCTCAGAGGACAGAGCAAAAGGCCGTGCATCCGCGACGG	1666
QY	360	aIProSerVaJProSerVaJGlnProSerLeuLeuGluAspSerProLeuVaJGlnAlaI	380
Db	1667	TGCCAGAGGTGCCAGCGGAGCGCGTGTGATGAGAGACAGCCCGCTGGTACAGGGCG	1726
QY	380	eUsErGlyLeuHnIstEusErAlaApAaGleuGluAspSerAsnLYsLeuLYsArgSerP	400
Db	1727	TCAGTGGGCTGCACTGTCTCCGACAGAGCTGGAAAGACAGAAATAGCTCAAGGGTCTC	1786
QY	400	heSerLeuAspIleLYsSerVaJSerTYsSerAlaSerMeCAlaIaSerLeuHnIagIYp	420
Db	1787	TCTCTCTGGATTCAAATCATAGTTCAATTCATTCAGCCAGATGGAGCATCTTACATGGCT	1846
QY	420	heSerSerSerGluAspAlaLeuGluTYTYTYLYsProSerThrLeuAspGlyThra	440
Db	1847	TCTCCTCATCAGAAAGATGCTTGGAAATCTCAAACTTCCACTACTGTGATGGAGCA	1906
QY	440	anLYsLeuCYsGlnPheSerProVaJGlnIulEusErGluGlnThrProGluThrSerP	460
Db	1907	ACAAAGCTATGCCAGTTCTTCCTCTGTTCAGAACTATCGAGAGAGCTCCGAAACCAATC	1966
QY	460	roAspLYsGluGluAlaSerIleProLYsLYsLeuGlnThrAlaAspProSerAspSerG	480
Db	1967	CTGATTAAGAGAGAGACCGACATCCCAAGAAAGCTCAGACCGCGAGCTTCAAGACGCC	2026
QY	480	IuSerLYsArgLeuHnIstSerVaJArgThrSerSerSerGlyThrAlaGlnaYsSerLeuI	500
Db	2027	AGAGCAACCGATGGATTCGGTCAGAAACACACAGACAGAGGAGCGCCAGAGGTCCTTT	2086
QY	500	eUsErProLeuHnIstArgSerGlySerVaJGluAspAsnTYHnIstHsErPheLeuPheG	520
Db	2087	TATCTCCACTGATGAGATGGAGCGTGGAGAGCAATTACACACCAAGCTTCTTTGCG	2146
QY	520	IyLeuSerThsErGlnGlnHnIstEuthrLYsSerAlaGlyLeuGlyLeuLYsGlyTYpH	540
Db	2147	GCTTTCCACACGACAGACAGCACTCAAGAAATGCTGGCGCTGGGCTTAAAGGGCTGCG	2206
QY	540	IstSerAspAlaLeuAlaProGlnThrSerThrProSerLeuThsErSerThrTYrPheA	560
Db	2207	ACTGGGATATCTGGCCCCCAAGACTCTTACCTTCTCTCAACAGCGCTGTATTTTG	2266
QY	560	IaThGluSerSerHnIstHsErTYsSerAlaSerAlaIaIeTYsGlyGlySerAlaSerTYs	580
Db	2267	CCACAGAGTCCATCACTTCTACTGTGCTCAGCCATTTAGAGAGGAGCTCCAGTTACT	2326
QY	580	eAlaIaTYsErCYsErGlnLeuPProThrCYsGlyAspGlnVaLYsErVaIaArga	600
Db	2327	CTGCTTAAGCTGCGACGCGCTCCCACTTGGGAGACCAAGTCAATCTGTGGCGAGCG	2386
QY	600	rgGlnLYsProSerAspArgAlaAspSerArgArgSerTYpHnIstGluIuSerProPheG	620
Db	2387	GGCAAAAGCCAAAGTACAGAGCTGACTCGCGCGAGAGCTGGCATTAAGAGACCCCTTTG	2446
QY	620	IuLYsGlnPheLYsArgArgSerCYsGlnMetGluPheGlyGluSerIleWetsErGluA	640
Db	2447	AAAGCAAGTTTAAAGCAGAAAGCTCCAAATGTGAATTTGGAGAGAGCTCATGTCTCAGGA	2506

QY	660	spargserArgGluGluuengLyysVa1c1ysserGlnserSerPheSerGlyserMetG	660
Db	2507	ACAGGTCCGGGAAAGAGCTGGGAGAAAGTGGCAGCTTACGCTTTTGGGCGAGCATGG	2566
QY	660	lu1e1e1eGluValSer	665
Db	2567	AAATCATTGAGGTCTCC	2583
RESULT 28			
ID	AAF30479	standard; cDNA; 2118 BP.	
XX	AAF30479;		
XX	29-MAY-2001	(first entry)	
DE	Human protein phosphatase and kinase protein-4 cDNA 1234795CBL.		
XX	Protein phosphatase and kinase protein; PPHK-4; human;		
KW	gastrointestinal disorder; immune system disorder; neurological disorder;		
KW	cell proliferative disorder; cancer; diagnosis; therapy; ss.		
OS	Homo sapiens.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	25..2025	
FT		/*tag= a	
FT	misc_feature	163..222	
FT		/*tag= b	
FT	misc_feature	/note= "unique fragment"	
FT		1081..1140	
FT		/*tag= c	
FT		/note= "unique fragment"	
PN	WO200120004-A2.		
XX			
PD	22-MAR-2001.		
XX			
XX	14-SEP-2000; 2000WO-US025515.		
PF			
XX	15-SEP-1999; 99US-0154141P.		
PR			
XX			
PA	(INCY-) INCYTE GENOMICS INC.		
PI	Yue H, Tang YT, Bandman O, Hillman JL, Baughn MR, Azimzai Y,		
PI	Lu DAM;		
XX			
DR	WPI: 2001-244811/25.		
XX	P-PSDB; AAB20325.		
PT	Novel human protein phosphatase and kinase proteins for diagnosis,		
PT	treatment and prevention of gastrointestinal, immune system, neurological		
PT	and cell proliferative disorders.		
XX			
PS	Claim 5; Page 97-98; 103pp; English.		
XX			
CC	The present sequence is that of cDNA encoding novel human protein		
CC	phosphatase and kinase protein PPHK-4 (see AAB20325). The cDNA was		
CC	initially identified in Incyte Clone ID No. 1234795CBL, from a foetal		
CC	tissue library. Tissues that express PPHK-4 (as a fraction of total		
CC	lung tissue expressing PPHK-4) include gastrointestinal (0.385),		
CC	cardiovascular (0.154), nervous (0.154) and reproductive (0.154).		
CC	Diseases or conditions associated with tissues expressing PPHK-4 (as a		
CC	fraction of total tissues expressing PPHK-4) include cancer (0.692),		
CC	inflammation or trauma (0.308) and cell proliferation (0.231). The		
CC	encoded protein shows homology to mouse neuronal tyrosine threonine		
CC	phosphatase 1. The invention provides human PPHK-1 to -11 polypeptides		
CC	(see AAB20322-32) and polynucleotides (see AAF30476-86). It also provides		
CC	expression vectors, host cells, antibodies, agonists and antagonists, as		
CC	well as methods for diagnosing, treating or preventing disorders		
CC	associated with expression of PPHK, including gastrointestinal		
CC	disorders, immune system disorders, neurological disorders and cell		
CC	proliferative disorders, including cancer		

XX Sequence 2118 BP; 550 A; 552 C; 506 G; 510 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No.: 0
Score: 439.00
Percent Similarity: 98.96%
Best Local Similarity: 98.96%
Query Match: 66.02%
DB: 4
Length: 2118
Matches: 663
Conservative: 0
Mismatch: 2
Indels: 7
Gaps: 0
US-10-029-345a-109 (1-665) x AAF30479 (1-2118)
QY 1 MetAlaHisGluMetIleGlyThrGlnIleValThrGluArgLeuValAlaLeuLeuGlu 20
DB 25 ATGGCCCATGAGTATGATGGAACTCAAAATTTGTAAGTGAAGTGGTGGCTCTGCTGGA 84
QY 21 SerGlyThrGluValLeuLeuIleAspSerArgProPheValGluTyrAsnThrSer 40
DB 85 AGTGAACCGAAAGAGTGTGCTAATGATGACCGGCATTTGTGGAATACATATCATCC 144
QY 41 HisIleLeuGluAlaIleAsnIleAsnGlySerIleLeuMetLysArgLeuGlnGln 60
DB 145 CACATTTTGGAAAGCCATTAATATCACTCTCCAAAGCTTAAGAAAGGTTGCAACAG 204
QY 61 AspLysValLeuIleThrGluLeuIleGlnHisSerAlaLysHisLysValAspIleAsp 80
DB 205 GACAAAGTGTATATACAGAGCTCATCCAGCATTACGCAACATTAAGTTGATGATGAT 264
QY 81 CysSerGlnLysValValValTyrAspGlnSerSerGlnAspValAlaSerLeuSerSer 100
DB 265 TGGAGTCAGAAAGGTGTAGTTAAGATCAAGCTCCCAAGATGTTGCTCTCTCTTCA 324
QY 101 AspCysPheLeuThrValLeuLeuGluLysLeuGluLysSerPheAsnSerValHisLeu 120
DB 325 GACTGTTTCTCATCTACTCTCGGATAACTGGAGAAAGAGCTTCAACTCTGTCACTG 384
QY 121 LeuAlaGlyGlyPheAlaGluPheSerArgCysPheProGlyLeuGlyGlnGlyLysSer 140
DB 385 CTTCGAGGTGGGTTTGCTGAGTCTCTGTTTCCCTGGCTCTGTCAAGAAATCC 444
QY 141 ThrLeuValProThrCysIleSerGlnProCysLeuProValAlaAsnIleGlyProThr 160
DB 445 ACTCTAAGTCCCACTGCACTTCTCGCTTCTTAACTGCTTCAACCAATGGGCAACC 504
QY 161 ArgIleLeuProAsnLeuThrLeuGlyCysGlnArgAspValLeuAsnLysLeuLeu 180
DB 505 CCAATTTCTTCCCAATCTTATCTTGGCTGCCAGCAGATGTCTCAACAAAGAGCTGAT- 563
QY 181 -GlnGlnAsnGlyIleGlyTyrValLeuAsnAlaSerTyr-ThrCysProLysProAsp 200
DB 564 GCAGCAGAAATGGGATTTGGTTATGTGTAAATGCCAGCA-TACGTGTCCAAAGCTGACT 622
QY 200 HeIleProGluSerHisPheLeuArgValProValAsnAspSerPheCysGluLysIle 220
DB 623 TTATCCCCCAGTCTCATTTCTCGCGTGCCTGTGAATGACAGCTTTTGTGAAAAATTT 682
QY 220 euProThrLeuAspLysSerValAspPheIleGluLysAlaLysAlaSerAsnGlyCys 240
DB 683 TCCCGTGTGGTGAACAAATCAAGATTTTCATTGAAAGCAAAAGCAACCTCCATGATG 742
QY 240 alLeuValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIle 260
DB 743 TTCTAGTCACTGTTAGCTGGGATCTCCCGCTCCGCCACATGCTATCGCTACATCA 802
QY 260 eLysArgMetAspMetSerLeuAspGluAlaTyrArgPheValLysGluLysArgProT 280
DB 803 TGAAGAGATGACATGCTTAACTTGAATGAAGCTTACAGATTTGTGAAAGAAAGAACTTA 862
QY 280 hrIleSerProAsnPheAsnPheLeuGlyGlnLeuLeuAspTyrGluLysLysIleLys 300
DB 863 CTATATCTCCAAACTTCAATTTTCTGGGCCAATCTCTGAACTATGAGAAAGATTAAGA 922

QY 300 snGlnThrGlyAlaSerGlyProLysSerLysLeuLysLeuLeuHisLeuGluLysPro 320
DB 923 ACCAGACTGAGACATCAGGCGCAAAAGCAAACTAACTGTGTGACCTGGAGAAAGCCAA 982
QY 320 snGluProValProAlaValSerGluGlyGlnLysSerGluThrProLeuSerProp 340
DB 983 ATGAACCTGTCTCTGTCTGTCTGAGAGGGTGGACAAAGAAAGCAAGCGCCCTCATTCAC 1042
QY 340 rOCyAlaAspSerAlaThrSerGluAlaAlaGlyGlnArgProValHisProAlaSer 360
DB 1043 CCTGTGGCCGACTTCTCACTTCAGAGCAGCAGCAAGCAAGGCCCTGTGATCCCGCAGCG 1102
QY 360 alProSerValProSerValGlnProSerLeuLeuGluAspSerProLeuValGlnAla 380
DB 1103 TCCCAAGCTGCCAGCTGAGCCGTGCTGTTAAGAGACAGCCCGCTGGTACAGGCGC 1162
QY 380 euSerGlyLeuHisLeuSerAlaAspArgLeuGluAspSerAsnLysLeuLysArgSer 400
DB 1163 TCAGTGGGCTGACCTGTCCGAGACAGCTGGAAAGACAGCAATAGCTCAAGCTTCTCT 1222
QY 400 heSerLeuAspIleLysSerValSerTyrSerAlaSerMetAlaAlaSerLeuHisGly 420
DB 1223 TCTCTGTGATATCAATCAAGTTTCATATTCAGCCAGCATGCGACATCTTATCAGCT 1282
QY 420 heSerSerSerGluAspAlaLeuGluTyrTyrLysProSerThrThrLeuAspGlyThr 440
DB 1283 TCTCTCTCATCAAGAGTCTTGGAAATCTCAAACTTCCACTACTGTGATGGAGACA 1342
QY 440 snLysLeuCysGlnPheSerProValGlnGluLeuSerGluGlnThrProGluThrSer 459
DB 1343 ACAAGCTATGCAAGTCTCCCTGTTCAAGAACTATCGAGACAGACTCCGAAACAGATT 1402
QY 460 -ProAspLysGlnGluAlaSerIleProLysLysLeuGlnThrAlaArgProSerAspSe 479
DB 1403 CCTGATTAAGAGAGAGCAGACATCCCAAGAACTGTGAGCCCGGCTTCAAGACAG 1462
QY 479 rGlnSerLysArgLeuHisSerValArgThrSerSerSerGlyThrAlaGlnArgSerLe 499
DB 1463 CCAAGCAGAGGATTTGATTTGGTCTGACAGACAGACAGAGTGGACCGCCAGAGTCTCT 1522
QY 499 uLeuSerProLeuHisArgSerGlySerValGluAspAsnTyrHisThrSerPheLeu 519
DB 1523 TTTATCTCACTGCATCGAAGTGGAGGTGAGAGACATTAACACACAGCTTCTTTT 1582
QY 519 eGlyLeuSerThrSerGlnGlnHisLeuThrLysSerAlaGlyLeuGly-LeuLysGly 539
DB 1583 CGGCTTTTCCACAGCCAGCAGCAGCTTCAGAAAGTGTGGCTGGCCCTTAAAGGCT 1642
QY 539 rPHisSerAspIleLeuAlaProGlnThrSerThrProSerLeuThrSerSerTyrP 559
DB 1643 GGCACCTGGATATCTTGGCCCCCAGACCTTACCCCTTCCCTGACAGCAGCTGATTT 1702
QY 559 heAlaThrGluSerSerHisPheTyrSerAlaSerAlaIleTyrGlyLysAlaSerT 579
DB 1703 TTGGCAGAGAGTCTTCACTTCACTGTGCTGACGCAATCTAGGAGGAGCAGTCCAGTT 1762
QY 579 rYrSerAlaTyrSerCysSerGlnLeuProThrCysGlyAspGlnValTyrSerValArg 599
DB 1763 ACTGTGCTTCAAGCTGAGCAGCTGCTCCACTTGGGAGACCAAGTTATTTGTGTGGCA 1822
QY 599 rGATGlnLysProSerAspArgAlaAspSerArgAspSerTrpHisGlnGluSerProp 619
DB 1823 GCGCGCAGAAAGCAAGTACAGAGCTACTCGCGCGAGCTGGCATGAAGAGAGCCCT 1882
QY 619 heGluLysGlnPheLysArgArgSerCysGlnMetGluPheGlyGluSerIleMetSerG 639
DB 1883 TTGAAAAGCAGTTTAAACGACAGAAAGCTGCCAAATGCAATTTGAGAGAGCATATGCA 1942
QY 639 LuAsnArgSerArgGluGlnLeuGlyLysValGlySerGlnSerSerPheSerGlySer 659
DB 1943 AGAACAGGTACGGGAGAGCTGGGAGAAAGTGGGAGTCACTGACTTTTTCGGGAGCA 2002
QY 659 etGluIleIleGluValSer 665


```

QY      359  rValProSerValProSerValGlnProSerLeuGluAspSerProLeuValGlnAl 379
DB      1546  CGTCCAGAGGTGCGCCAGCGCGCGTGTAGAGACAGCCCGGTGTACAGGC 1605
QY      379  aLeuSerGlyLeuHisLeuSerAlaAspArgLeuGluAspSerAsnLysLeuValAspSe 399
DB      1606  GCTCAGTGGGCTGACCTGTCTCCGACAGACGCTGGAAAGACCAATTAAGCTCAAGCCTTC 1665
QY      399  rPheSerLeuAspLysSerValSerTyrSerAlaSerMetAlaAlaSerLeuHisGly 419
DB      1666  CTCTCTCTGGAATATCAAAATCAGTTTCATATTCAGCCAGCATGGCAGCATCTTACATGG 1725
QY      419  yPheSerSerSerGluAspAlaLeuGluTyrTyrTyrProSerThrThrLeuAspGlyTh 439
DB      1726  CTCTCTCTCATCAAGAGATCTTTGGAATCTACAAACCTTCCACTACTCTGTGATGGAGAC 1785
QY      439  rAsnLysLeuGlyAspGlnPheSerProValGlnGluLeuSerGluGlnThrProGluThrSe 459
DB      1786  CAACAAGCTATGCGCAGTTCTCCCTGTTCAGGAACCTATCGAGCAGACTCCGAAACCAAG 1845
QY      459  rProAspLysGluGluAlaSerLysLeuLysLysLeuGlnThrAlaArgProSerAspSe 479
DB      1846  TCCGTAAAGAGAGAGACCCAGCATCCCAAGAAAGCTGACAGCCGACGCTTCAACAGAG 1905
QY      479  rGlnSerTyrArgLeuHisSerValArgThrSerSerSerGlyThrAlaGlnArgSerLe 499
DB      1906  CCAAGAGACGATGCTGATTCGCTCAGAACACACAGACAGTGGACCGCCAGAGGTCTCT 1965
QY      499  uLeuSerProLeuHisArgSerGlySerValGluAspAsnTyrHisThrSerPheLeuTh 519
DB      1966  TTATATCTCCACTGATCGAAGTGGAGCGTGGAGGACCAATTACACACAGCTTCTCTTT 2025
QY      519  eGlyLeuSerThrSerGlnGlnHisLeuThrLysSerAlaGlyLeuGlyLeuLysGlyTyr 539
DB      2026  CGGCTTTTCCACAGCAGCAGACGACCTTCAGAAAGTGTGCTGGCGCTTAAAGGCTCG 2085
QY      539  pHisSerAspLysLeuAlaProGlnThrSerThrProSerLeuThrSerSerTriPyrPh 559
DB      2086  GCACCTCGAATATCTTGCCGCCCCAGACCTCTACCCCTTCCCTGACACAGACCTGTATTT 2145
QY      559  eAlaThrGluSerSerHisPheTyrSerAlaSerAlaLeuTyrGlyGlySerAlaSerTyr 579
DB      2146  TGGCACAAGAGTCTCTCACTTCTACTCTGCTCAGCAGCATCTACGAGGAGCAGTCCAGTTA 2205
QY      579  rSerAlaTyrSerCysSerGlnLeuProThrCysGlyAspGlnValTyrSerValArgAr 599
DB      2206  CTCTGCTCTACAGTGCAGCCAGCTGCCACTTGGCGAGACCAAGCTTATTTGTGCGCAG 2265
QY      599  gArgGlnLysProSerAspArgAlaAspSerArgArgSerTPHISGluGluSerProPh 619
DB      2266  GCGGCGAGAGCCAAAGTACAGAGCTGACTCGCGCGAGAGCTGCATGAAGAGACCCCTT 2325
QY      619  eGluLysGlnPheLysArgArgSerCysGlnMetGluPheGlyGluSerTyrLeuSerGly 639
DB      2326  TTAATAAGCAGTTTAAACGCAAGAGCTGCCAAATTTGAATTTGAGAGAGCATCATGTGCA 2385
QY      639  uAsnArgSerArgGluGluLeuGlyLysValGlySerGlnSerSerPheSerGlySerMet 659
DB      2386  GAAACAGGTACCGGAGAGAGCTGGGAAAGTGGGACAGTCACTTCTTGGGAGAGAT 2445
QY      659  tGluLeuLeuGluValSer 665
DB      2446  GGAATTCATGAGGTCTCC 2464

```

RESULT 30
ACCG60560
ID ACC60560 standard; cDNA; 2756 BP.

ACCG60560;
XX
XX
DT 19-JUN-2003 (first entry)
XX

```

DE Polynucleotide relating to the invention SEQ ID NO: 113.
XX
XX Gene; ss; antiproliferative; hepatotropic; nephrotropic; antiarthritic;
KM antipsoriatic; cardiant; cytosaric; gene therapy; liver disease;
KM proliferative disorder; renal failure; cardiovascular disorder;
KM immunological disorder; arthritis; psoriasis; congenital heart defect;
KM congestive heart failure; cancer; human; mouse; yeast; rat; fruitfly.
XX
OS Mus musculus.
XX
XX WO200257460-A2.
XX
XX 25-JUL-2002.
XX
XX 20-DEC-2001; 2001WO-US050459.
XX
XX 20-DEC-2000; 2000US-0256868P.
PR 30-MAR-2001; 2001US-0280186P.
PR 01-MAY-2001; 2001US-0287735P.
PR 05-JUN-2001; 2001US-0295848P.
PR 25-JUN-2001; 2001US-0300465P.
XX
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
XX Jackson DG, Feder J, Nelson T, Mintier G, Ramanathan C, Lee L,
PI Siemers N, Bol D, Schieven G, Finger J, Todderud CG, Bassolino D,
PI Krytek S, Mcatee P, Suchard S, Banae D;
XX
XX WPI; 2002-599721/64.
DR P-PSDB; ABR52385.
XX
XX Novel polynucleotides encoding human phosphatase polypeptides useful in
PT the prevention or treatment of e.g. proliferative and cardiovascular
PT disorders.
XX
XX Disclosure; Fig 16; 801bp; English.
XX
XX The invention relates to a novel isolated nucleic acid comprising a
CC polynucleotide having a nucleotide sequence selected from 40
CC polynucleotides fully defined in the specification. The polynucleotide of
CC the invention has antiproliferative, hepatotropic, nephrotropic,
CC antarthritic, antipsoriatic, cardiant, and cytosaric activity. The
CC polynucleotide may have a use in gene therapy. A polynucleotide or
CC polypeptide of the invention is useful for preventing, treating or
CC ameliorating a medical condition, e.g. a proliferative disorder. They are
CC also useful for treating e.g. liver disease, renal failure, immunological
CC disorders including arthritis and psoriasis, cardiovascular disorders
CC such as congenital heart defects and congestive heart failure, and
CC cancer. A method of the invention is useful for diagnosing a pathological
CC condition or susceptibility to a condition in a subject. The present
CC sequence is used in the exemplification of the invention
XX
XX
SQ Sequence 2756 BP; 714 A; 675 C; 664 G; 703 T; 0 U; 0 Other;
XX
XX
Alignment Scores:
Pred. No.: 1 49e-52 Length: 2756
Score: 64.00 Matches: 93
Percent Similarity: 97.89% Conservative: 0
Best Local Similarity: 97.89% Mismatches: 1
Query Match: 9.62% Indels: 2
DB: 6 Gaps: 0
XX
US-10-029-345A-109 (1-665) x ACC60560 (1-2756)
QY 194 ThrCysProLysProAspPheLeuProGluSerHisPheLeuArgValProValAsnAsp 213
DB 948 ACCTGTCGAAGCCTGATCTTCAATACCGAATCTCACTTCTGCGAGAGCTGTGAATGAC 1007
QY 214 SerPheCysGluLysLeuPheProTyrPheLeuAspLysSerValAspPheLeuGlyVal 233
DB 1008 AGCTTTGTGAAGAAATCTTACCATGTGGAGCAAAAGTGTGATTTCAATTGGAAGAACA 1067
QY 234 LysAlaSerAsnGlyCysVal-LeuValHisCysLeuAlaGlyLysSerArgSerAlaThr 253

```

```

Db      1068 AAAGCTCCATGCGTGTGCTATC-CAGTCTAGCTGGGATCTCGCCGCCAC 1126
Qy      253 rllalaiallaaTyrllleWeclysaRgWecAapWctSerleuAepGlualaTyrAph 273
Db      1127 TATTTGCTATGCTCTACATCATGAGAGATGACACAGTCTCTAGAGAGGCTTACAGATT 1186
Qy      273 eVallysGlylysaRgProThrllleSerProAenPheAenPhe 287
Db      1187 TGTGAAAGAAAAGACCTACTATATCTCGAAATTTAATTTT 1229

RESULT 31
ADE07547
ID      ADE07547 standard; DNA; 1917 BP.
AC      ADE07547;
XX      29-JAN-2004 (first entry)
DE      Novel coding sequence (useful for identifying genetic disorders) #613.
XX      novel gene; novel protein; tissue marker; molecular weight marker;
XX      chromosome marker; genetic disorder; gene; de.
XX      Unidentified.
XX      MO2003054152-A2.
XX      03-JUL-2003.
XX      10-DEC-2002; 2002MO-US039555.
XX      10-DEC-2001; 2001US-0339739P.
XX      PR 11-DEC-2001; 2001US-0339453P.
XX      PR 14-MAR-2002; 2002US-0365091P.
XX      PR 14-MAR-2002; 2002US-0365384P.
XX      PR 12-APR-2002; 2002US-0372381P.
XX      PR 12-APR-2002; 2002US-0372615P.
XX      PR 22-APR-2002; 2002US-00128558.
XX      PR 24-APR-2002; 2002US-0376045P.
XX      PA
XX      (HYSE-) HYSEQ INC.
XX      Tang YT, Asundi V, Goodrich RW, Ren F, Zhang J, Zhao QA, Wang J;
XX      Ghosh W, Xue AJ, Wehrman T, Weng G, Zhou P, Drmanac RT, Wang Z;
XX      Ma Y, Wang D, Chen R, Xu C, Boyle BJ;
XX      WPI; 2003-569235/53.
XX      DR P-PSDB; ADE08458.
XX      New polynucleotides, useful for expressing recombinant proteins for
XX      PT analysis, characterization or therapeutic use, or as markers for tissues
XX      PT in which the corresponding protein is preferentially expressed.
XX      Claim 1; SEQ ID NO 613; 1177bp; English.
XX      The invention comprises the amino acid and coding sequences of novel
XX      CC proteins. The DNA and protein sequences of the invention are useful as:
XX      CC markers for tissues in which the corresponding protein is preferentially
XX      CC expressed; as molecular weight markers on gels; as chromosome markers or
XX      CC tags; to identify chromosomes or to map related gene positions; and to
XX      CC compare with endogenous DNA sequences in patients to identify potential
XX      CC genetic disorders. The present DNA sequence represents a gene of the
XX      CC invention.
XX      SQ Sequence 1917 BP; 305 A; 724 C; 613 G; 275 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2.86e-09 Length: 1917
Score: 20.00 Matches: 20
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.01% Indels: 0

```

```

DB: 9 Gaps: 0
US-10-029-345A-109 (1-665) x ADE07547 (1-1917)
Qy      242 VALHISQYsleuAlGlyllleSerArgSerAlaThrllleAlaTyrllleWecly 261
Db      733 GTCCACTGCTGCTGGCGATCTCCGCTCTGCACCATGCGCATCATGAG 792

RESULT 32
ACA64956
ID      ACA64956 standard; DNA; 2377 BP.
AC      ACA64956;
XX      27-JUN-2003 (first entry)
DE      Human protein tyrosine phosphatase DNA corresponding to U27193.
XX      Human; chronic inflammatory joint disease; infection; tumour;
XX      antiinflammatory; cytosratic; antirheptic; antirheumatic;
XX      immunosuppressive; gene therapy; etiological pathogenicity; de.
XX      Homo sapiens.
XX      DE10127572-A1.
XX      PD 05-DEC-2002.
XX      PF 30-MAY-2001; 2001DE-01027572.
XX      PR 30-MAY-2001; 2001DE-01027572.
XX      PA (PATH-) PATHOARRAY GMBH.
XX      PI Haepul T, Ungethuem U, Blaess S;
XX      WPI; 2003-240797/24.
XX      DR Reagents for diagnosis, study and therapy of chronic inflammatory joint
XX      PT and other diseases, comprises any of many specified genes or derived
XX      PT proteins.
XX      Claim 1; Page; 12pp; German.
XX      This invention describes a novel reagent for diagnosis, molecular
XX      CC definition and therapy of chronic inflammatory joint diseases, and other
XX      CC inflammatory disorders, infective or tumour diseases in humans. The
XX      CC products of the invention have antinflammatory, cytosratic,
XX      CC antirheptic, antirheumatic and immunosuppressive activity and can be
XX      CC used for gene therapy. The reagent of the invention and any proteins and
XX      CC antibodies derived from it, are used (i) for analysing tissue and blood
XX      CC samples for medical diagnosis; (ii) for diagnosis and characterisation of
XX      CC chronic joint diseases, on the basis of molecular characterisation, and
XX      CC determining the etiological pathogenicity principle of as yet
XX      CC uncharacterised inflammatory diseases, also monitoring progression and/or
XX      CC treatment of disease, and optimisation of therapy and (iii) for
XX      CC developing treatments for inflammatory diseases, particularly of joints,
XX      CC infections and tumours. ACA64801-ACA64965 represent human polynucleotides
XX      CC used in the method of the invention.
XX      SQ Sequence 2377 BP; 397 A; 862 C; 736 G; 382 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 3.52e-09 Length: 2377
Score: 20.00 Matches: 20
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.01% Indels: 0
DB: 7 Gaps: 0
US-10-029-345A-109 (1-665) x ACA64956 (1-2377)
Qy      242 VALHISQYsleuAlGlyllleSerArgSerAlaThrllleAlaTyrllleWecly 261

```

Db 864 GTCCAGTGTGGCTGGCATCTCCGCTCTGCCACCATGCCATCGCCTACATCATGAA 923

RESULT 33

ABX10760 standard; DNA; 2377 BP.

ABX10760;

10-MAY-2003 (first entry)

Human dual specific phosphatase 8 DNA.

Human; dual specific phosphatase 8; gene; ds; infection; inflammation;

tumour formation; cytostatic; antiinflammatory.

Homo sapiens.

Location/Qualifiers

135..2012

/+tag= a

/product= "Human dual specific phosphatase 8"

US6482644-BI.

19-NOV-2002.

01-AUG-2001; 2001US-00920668.

01-AUG-2001; 2001US-00920668.

(ISIS-) ISIS PHARM INC.

Cowser LM;

WPI; 2003-298140/29.

P-PSDB; AEG73440.

New antisense compound targeted to a nucleic acid encoding human dual specific phosphatase 8, for modulating gene expression and treating diseases associated with expression of the phosphatase in humans.

Claim 1; Col 49-54; 36pp; English.

The invention relates to a compound targeted to the coding region of a nucleic acid encoding human dual specific phosphatase 8, where the compound specifically hybridises with the region and inhibits the expression of human dual specific phosphatase 8. The compound is useful for inhibiting the expression of human dual specific phosphatase 8 in cells or tissues, and for treating an animal, particularly a human, suspected of having or being prone to a disease or condition associated with expression of dual specific phosphatase 8. The compound is useful for diagnosis, therapeutics and as a research reagent, e.g. to prevent or delay infection, inflammation or tumour formation, and to distinguish between functions of various members of a biological pathway. This sequence represents DNA encoding human dual specific phosphatase 8

Sequence 2377 BP; 397 A; 862 C; 736 G; 382 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	3.52e-09	Length:	2377
Score:	20.00	Matches:	20
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.01%	Indels:	0
DB:	7	Gaps:	0

US-10-029-345A-109 (1-665) x ABX10760 (1-2377)

Qy 242 ValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIleMetLys 261
Db 864 GTCCAGTGTGGCTGGCATCTCCGCTCTGCCACCATGCCATCGCCTACATCATGAA 923

RESULT 34

AAT86757 standard; cDNA; 2415 BP.

AAT86757;

15-DEC-1997 (first entry)

cDNA of the M3/6 gene.

murine; phosphatase; M3/6; dual specificity; Threonine; Tyrosine;

inactivate; mitogen activated protein kinase; MAP-K; cdc25 PRP-yeast;

trinucleotide repeat; repeat expansion; neurodegenerative; proliferative;

diagnosis; tumour; lung; brain; chromosomal deletion; ss.

Mus sp.

WO9706245-A1.

20-FEB-1997.

05-AUG-1996; 96WO-GB001906.

04-AUG-1995; 95GB-00016059.

(MEDI-) MEDICAL RES CONNCTL.

Davies KE, Theodosiou A;

WPI; 1997-154253/14.

Murine mitogen activated protein (MAP) kinase phosphatase M3/6 - is a suspected dual specificity Thr-Tyr phosphatase, useful for diagnosing and treating neuro-degenerative or proliferative diseases e.g. tumours.

Claim 1; Fig 1; 51pp; English.

This cDNA is said to encode a murine phosphatase designated M3/6 (see also AAT86758). M3/6 is a suspected dual specificity Threonine-Tyrosine phosphatase, capable of inactivating mitogen activated protein (MAP) kinase. The M3/6 protein product shows high homology to the cdc25 PRP of yeast at residues 29-49 and 117-136. The gene also contains a complex triplet distal to the catalytic domain which is translated into the protein. This domain comprises a run of 4 serine residues which in turn is followed by a further run comprising 23 serine residues which is interrupted near the N-terminal section by a single asparagine. This makes the phosphatase gene a candidate for a human disease caused by repeat expansion or mutation. M3/6 is expressed highly in the brain and may have utility in investigating signal transduction mechanisms in brain and muscle. The M3/6 and Hb5 (a human homologue) genes may be responsible, if mutated, for various neurodegenerative or proliferative diseases, and may therefore be used for the diagnosis of such diseases, e.g. tumours, especially lung or brain tumours, associated with deletion of the chromosomal region 11p15.5. The polypeptides can be used to screen for inhibitors to treat these diseases

Sequence 2415 BP; 491 A; 764 C; 678 G; 465 T; 0 U; 17 Other;

Alignment Scores:

Pred. No.:	3.57e-09	Length:	2415
Score:	20.00	Matches:	20
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	3.01%	Indels:	0
DB:	2	Gaps:	0

US-10-029-345A-109 (1-665) x AAT86757 (1-2415)

Qy 242 ValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIleMetLys 261
Db 825 GTTCAGTGTGGCTGGCATCTCCGCTCTGCCACCATGCCATCGCCTACATCATGAA 884

RESULT 35

AAAT86758
 ID AAT86758 standard; cDNA, 2453 BP.
 AC AAT86758;
 XX
 XX 15-DEC-1997 (first entry)
 DE cDNA of the M3/6 gene.
 XX
 XX murine; phosphatase; M3/6; dual specificity; Threonine; Tyrosine;
 KW inactive; mitogen activated protein kinase; MAP-K; cdc25 pTP; yeast;
 KM trinucleotide repeat; repeat expansion; neurodegenerative; proliferative;
 KM diagnosis; tumour; lung; brain; chromosomal deletion; ss.
 XX
 OS Mus sp.
 XX
 XX Key Location/Qualifiers
 FH 98..2089
 FT /*tag= a
 FT /product= "dual-specific_phosphatase"
 XX
 XX MO9706245-A1.
 XX
 XX 20-FEB-1997.
 XX
 XX PD 05-AUG-1996; 96MO-GB001906.
 XX
 XX PR 04-AUG-1995; 95GB-00016059.
 XX
 XX PA (MED1-) MEDICAL RES COUNCIL.
 XX
 XX PI Davies KE, Theodorou A;
 XX
 XX DR MPI, 1997-154253/14.
 XX
 XX DR P-PSDB; AAM29150.
 XX
 XX PT Murine mitogen activated protein (MAP) kinase phosphatase M3/6 - is a
 PT suspected dual specificity Thr-Tyr phosphatase, useful for diagnosing and
 PT treating neuro-degenerative or proliferative diseases e.g. tumours.
 XX
 XX claim 21; Fig 2; 51pp; English.
 XX
 XX PS This cDNA encodes a murine phosphatase designated M3/6 (see also
 CC AAT86757). M3/6 is a suspected dual specificity Threonine-Tyrosine
 CC phosphatase, capable of inactivating mitogen activated protein (MAP)
 CC kinase. The M3/6 protein product shows high homology to the cdc25 pTP of
 CC yeast at residues 29-49 and 117-136. The gene also contains a complex
 CC triplet distal to the catalytic domain which is translated into the
 CC protein. This domain comprises a run of 4 serine residues which in turn
 CC is followed by a further run comprising 23 serine residues which is
 CC interrupted near the N-terminal section by a single asparagine. This
 CC makes the phosphatase gene a candidate for a human disease caused by
 CC repeat expansion or mutation. M3/6 is expressed highly in the brain and
 CC may have utility in investigating signal transduction mechanisms in brain
 CC and muscle. The M3/6 and Hb5 (a human homologue) genes may be
 CC responsible, if mutated, for various neurodegenerative or proliferative
 CC diseases, and may therefore be used for the diagnosis of such diseases,
 CC e.g. tumours, especially lung or brain tumours, associated with deletion
 CC of the chromosomal region 11p15.5. The polypeptides can be used to screen
 CC for inhibitors to treat these diseases
 XX
 SO Sequence 2453 BP; 493 A; 784 C; 705 G; 471 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 3.62e-09 Length: 2453
 Score: 20.00 Matches: 20
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 3.01% Indels: 0
 DB: 2 Gaps: 0

US-10-029-345A-109 (1-665) X AAT86758 (1-2453)

OY 242 VALHSCYELenAAGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIleMetLys 261
 |||||
 DB 827 GTTCACTGCTGTGGCTGCATCTCTGCTCTGCACCATTCGCCATCCGTACATCATGAAA 886
 RESULT 36
 ID AAD41236
 XX AAD41236 standard; cDNA, 2453 BP.
 XX
 AC AAD41236;
 XX
 XX 30-OCT-2002 (first entry)
 DT
 XX
 XX Murine neuronal tyrosine/threonine phosphatase (NTPP1) cDNA.
 DE
 XX
 XX Murine; targeting construct; neuronal tyrosine/threonine phosphatase;
 KW NTPP1 gene; transgenic mouse; pharmacological therapy; genetic disease;
 KW neurological; neuropsychological; psychotic illness; ss.
 XX
 OS Mus musculus.
 XX
 XX Key Location/Qualifiers
 FH 1..142
 FT /*tag= a
 FT /note= "Sequence flanking Neo insert in targeting
 FT construct"
 FT misc_feature 143..304
 FT /*tag= b
 FT /note= "Deleted in targeting construct"
 FT misc_feature 305..328
 FT /*tag= c
 FT /note= "Sequence flanking Neo insert in targeting
 FT construct"
 XX
 XX PN WO200245497-A2.
 XX
 XX PD 13-JUN-2002.
 XX
 XX PP 05-DEC-2001; 2001MO-US046868.
 XX
 XX PR 06-DEC-2000; 2000US-0251802P.
 XX
 XX PR 04-DEC-2001; 2001US-00005858.
 XX
 XX PA (DELTA-) DELTAGEN INC.
 XX
 XX PI Allen KD;
 XX
 XX DR MPI, 2002-537527/57.
 XX
 XX PT Novel targeting construct for producing a transgenic mouse useful as a
 PT disease model, has nucleotide sequences homologous to a neuronal
 PT tyrosine/threonine phosphatase gene and a selectable marker.
 XX
 XX PS Example 1; Fig 2A; 48pp; English.
 XX
 XX CC The invention relates to a targeting construct comprising a first
 CC polynucleotide sequence homologous to at least a first portion of
 CC neuronal tyrosine/threonine phosphatase (NTPP1) gene, a second
 CC polynucleotide sequence homologous to at least a second portion of the
 CC (NTPP1) gene and a selectable marker. The invention is useful for
 CC producing a transgenic mouse comprising a disruption in a NTPP1 gene.
 CC The invention is useful for identifying an agent that ameliorates a phenotype
 CC associated with a disruption in a NTPP1 gene. The invention is useful for
 CC relating a unique animal model for testing and developing new treatments
 CC relating to the behavioural phenotypes, which can be used to test the
 CC efficacy of proposed genetic and pharmacological therapies for human
 CC genetic disease, such as neurological, neuropsychological, or psychotic
 CC illnesses. The invention is utilized as models for diseases, disorders,
 CC or conditions associated with phenotypes relating to a disruption in a
 CC NTPP1 gene, to identify drugs, pharmaceuticals, drugs, therapies and
 CC interventions that are effective in treating disease. The present
 CC sequence is murine NTPP1 cDNA
 XX
 SO Sequence 2453 BP; 493 A; 784 C; 705 G; 471 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 3.62e-09
Score: 20.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 3.01%
Length: 2453
Matches: 20
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0
US-10-029-345A-109 (1-665) x AAD41236 (1-2453)
QY 242 ValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrlleMetlys 261
Db 827 GTTCACGTCTGCTGCGCATCTCTCGCTCGCCACCATTCGCATCATCATGAAA 886
RESULT 37
AAS31013
ID AAS31013 standard; cDNA; 2476 BP.
AC AAS31013;
XX
DT 04-DEC-2001 (first entry)
XX
DE Human diagnostic and therapeutic polynucleotide (DITHP) #28.
XX
KW Human; receptor; diagnostic; therapeutic; gene therapy; vaccine;
KW cell proliferative disorder; Crohn's disease; lymphoma; leukemia;
KW acquired immune deficiency syndrome; AIDS; autoimmune disorder;
KW respiratory disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200162927-A2.
PD 30-AUG-2001.
XX
PF 21-FEB-2001; 2001WO-US006059.
XX
PR 24-FEB-2000; 2000US-0184693P.
PR 24-FEB-2000; 2000US-0184697P.
PR 24-FEB-2000; 2000US-0184698P.
PR 24-FEB-2000; 2000US-0184768P.
PR 24-FEB-2000; 2000US-0184769P.
PR 24-FEB-2000; 2000US-0184770P.
PR 24-FEB-2000; 2000US-0184771P.
PR 24-FEB-2000; 2000US-0184772P.
PR 24-FEB-2000; 2000US-0184773P.
PR 24-FEB-2000; 2000US-0184774P.
PR 24-FEB-2000; 2000US-0184776P.
PR 24-FEB-2000; 2000US-0184777P.
PR 24-FEB-2000; 2000US-0184813P.
PR 24-FEB-2000; 2000US-0184813P.
PR 24-FEB-2000; 2000US-0184813P.
PR 24-FEB-2000; 2000US-0184841P.
PR 24-FEB-2000; 2000US-0185213P.
PR 24-FEB-2000; 2000US-0185216P.
PR 12-MAY-2000; 2000US-0203785P.
PR 15-MAY-2000; 2000US-0204226P.
PR 16-MAY-2000; 2000US-0204525P.
PR 16-MAY-2000; 2000US-0204821P.
PR 16-MAY-2000; 2000US-0204908P.
PR 16-MAY-2000; 2000US-0205232P.
PR 17-MAY-2000; 2000US-0204815P.
PR 17-MAY-2000; 2000US-0204863P.
PR 17-MAY-2000; 2000US-0205221P.
PR 17-MAY-2000; 2000US-0205285P.
PR 17-MAY-2000; 2000US-0205286P.
PR 17-MAY-2000; 2000US-0205287P.
PR 17-MAY-2000; 2000US-0205323P.
PR 17-MAY-2000; 2000US-0205324P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX

PI Panzer SR, Spiro PA, Banville SC, Shah P, Chalup MS, Chang SC;
PI Chen A, D'sa SA, Amshy S, Dahl CR, Dam TC, Daniels SE, Dufour GB;
PI Flores V, Fong WT, Greenwalt LB, Hillman JU, Jones AL, Liu TF;
PI Roseberry AM, Rosen BH, Russo PB, Stockdreyer TK, Daffo A;
PI Wright RJ, Yap PE, Yu JY, Bradley DL, Bratcher SR, Chen W;
PI Cohen HJ, Hodgson DM, Lincoln SE, Jackson S;
XX
DR MPI, 2001-502867/55.
DR P-PSDB; AAU19442.
XX
PT Polynucleotides encoding diagnostic and therapeutic proteins, e.g.
PT enzymes, hormones and receptors, useful in diagnostics and therapeutics.
XX
PS Claim 1; Page 310, 522pp; English.
XX
CC The invention relates to polynucleotides (I) encoding diagnostic and
CC therapeutic (DITHP) polypeptides (II), which include e.g. enzymes and
CC proteins involved in growth and development and receptors. (I) and (II)
CC may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate DITHP expression. For example, (I) and (II)
CC may be used to treat disorders associated with decreased polypeptide
CC expression by rectifying mutations or deletions in a patient's genome,
CC that affect the activity of the DITHPs, by expressing inactive proteins
CC or supplementing the patient's own production of them. (I) and (II) may
CC be used to treat diseases, for example, cell proliferative disorder,
CC Crohn's disease, acquired immune deficiency syndrome (AIDS), lymphoma,
CC leukaemia, autoimmune disorders, and respiratory disorders. Additionally,
CC (I) may be used to produce the DITHPs, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the protein. (I) and
CC its complementary sequences may also be used as DNA probes in diagnostic
CC assays to detect and quantitate the presence of similar nucleic acids in
CC samples, and therefore which patients may be in need of restorative
CC therapy. (II) may also be used as antigens in the production of
CC antibodies against DITHPs and in assays to identify modulators of DITHP
CC expression and activity. The anti-DITHP antibodies and antagonists may
CC also be used to down regulate expression and activity. The anti-DITHP
CC antibodies may also be used as diagnostic agents for detecting the
CC presence of DITHPs in samples (e.g. by enzyme linked immunosorbant assay
CC (ELISA)). AAS30986-AAS31136 represent human diagnostic and therapeutic
CC (DITHP) polynucleotides of the invention
XX
SQ Sequence 2476 BP; 432 A; 884 C; 762 G; 398 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 3.66e-09
Score: 20.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 3.01%
Length: 2476
Matches: 20
Conservative: 0
Mismatch: 0
Indels: 0
Gaps: 0
US-10-029-345A-109 (1-665) x AAS31013 (1-2476)
QY 242 ValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrlleMetlys 261
Db 933 GTTCACGTCTGCTGCGCATCTCTCGCTCGCCACCATTCGCATCATCATGAAA 992
RESULT 38
ABL13421
ID ABL13421 standard; cDNA; 2369 BP.
XX
AC ABL13421;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 34745.
XX
KW Drosophila; developmental biology; cell signaling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.

```

XX 27-SEP-2001.
PD ABL13420
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
XX
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE ) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
DR P-PSDB; ABB69317.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
XX Claim 1; SEQ ID NO 34745; 21bp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB161840-AB16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2369 BP; 716 A; 591 C; 557 G; 505 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 0.0285 Length: 2369
Score: 13.00 Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.95% Indels: 0
DB: Gaps: 0
XX
US-10-029-345A-109 (1-665) x ABL13421 (1-2369)
XX
QY 246 AAGAGYIIeSerARgSerAlaThrIIeAlaIIeAlaTyr 258
XX
DB 1125 GCGGGGATTTCGCGAGCGCCACCATTCGCATCGCCTAC 1163
XX
RESULT 39
AB13420
ID ABL13420 standard; cDNA; 18413 BP.
XX
AC ABL13420;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 34742.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
XX
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE ) PE CORP NY.

```

```

XX Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
DR P-PSDB; ABB69317.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions.
XX
XX Claim 1; SEQ ID NO 34742; 21bp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB161840-AB16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 18413 BP; 5536 A; 3801 C; 3766 G; 5310 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 0.202 Length: 18413
Score: 13.00 Matches: 13
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.95% Indels: 0
DB: Gaps: 0
XX
US-10-029-345A-109 (1-665) x ABL13420 (1-18413)
XX
QY 246 AAGAGYIIeSerARgSerAlaThrIIeAlaIIeAlaTyr 258
XX
DB 16169 GCGGGGATTTCGCGAGCGCCACCATTCGCATCGCCTAC 16207
XX
RESULT 40
ABV74143
ID ABV74143 standard; cDNA; 1435 BP.
XX
AC ABV74143;
XX
DT 23-JAN-2003 (first entry)
XX
DE Human dual specificity protein phosphatase-7-like protein cDNA.
XX
KW Dual-specificity protein phosphatase 7-like protein; DSPP7; human;
KW protein phosphatase; enzyme; cancer; asthma; obesity; diabetes;
KW cardiovascular disorder; chronic obstructive pulmonary disorder;
KW antitumor; antiparkinsonian; nootropic; neuroprotective; cardiac;
KW hypotensive; antiarrhythmic; antianginal; analgesic; antiallergic;
KW antiinflammatory; anorectic; antidiabetic; cranioillizer; antitonic;
KW antidepressant; dermatological; gene therapy; chromosome 3p21.1; gene;
KW ss.
XX
OS Homo sapiens.
XX
PN 179. .1435
XX
FT Key Location/Qualifiers
XX
FT CDS 179. .1435
XX
FT /tag=a
XX
FT /partial
XX
FT /product="DSPP7-like protein"
XX
FT /note="the CDS does not include a stop codon; this
FT region is also claimed in Claim 19"
XX
XX WO200277217-A2.
XX
PD 03-OCT-2002.
XX
PF 20-MAR-2002; 2002WO-EP003091.
XX

```


XX 23-MAR-2001; 2001US-0277954P.
 PR 28-JAN-2002; 2002US-0351377P.
 PA (FARB) BAYER AG.
 PI Zhu Z;
 XX WPI: 2003-029935/02.
 DR P-PSDB; ABP55026.
 XX
 PT New dual specificity protein phosphatase (DSPP)-7-like protein polypeptide
 PT and polynucleotide, useful for regulating DSPP-7-like protein activity for
 PT preventing, treating or ameliorating cancer or cardiovascular disorders.
 XX
 PS Claim 19; Fig 4; 132pp; English.
 XX
 CC The present sequence is that of a cDNA clone, the coding region of which
 CC is also claimed, encoding a novel human dual specificity protein
 CC phosphatase 7 (DSPP7)-like protein. The sequence is located on chromosome
 CC 3p21.1. The invention provides human DSPP7-like polypeptides and
 CC polynucleotides, and methods of using them for identifying reagents which
 CC regulate DSPP7-like protein or which bind to DSPP7-like gene products.
 CC These can be used in preventing, ameliorating or correcting dysfunctions
 CC or diseases associated with DSPP7-like protein dysfunction such as
 CC cancer, central nervous system disorders (e.g. mood disorders, anxiety,
 CC Parkinson's disease, Alzheimer's disease), asthma (including other
 CC allergic diseases such as allergic rhinitis or atopic dermatitis),
 CC cardiovascular disorders (e.g. myocardial infarction, hypertension,
 CC arrhythmias or angina pectoris), chronic obstructive pulmonary disease,
 CC obesity, or diabetes (all claimed). They can also be used to treat pain
 CC associated with the above disorders. The polynucleotides can also be used
 CC in diagnostic assays or in genetic testing, as probes or primers, in the
 CC production of DSPP7-like recombinant polypeptides, in gene therapy.
 CC Antisense oligonucleotides and ribozymes are used in a claimed method of
 CC reducing human DSPP7-like protein activity
 XX
 SQ Sequence 1435 BP; 252 A; 517 C; 452 G; 207 T; 0 U; 7 Other;
 Alignment Scores:
 Pred. No.: 0.171 Length: 1435
 Score: 12.00 Matches: 12
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.80% Indels: 0
 DB: 7 Gaps: 0
 US-10-029-345A-109 (1-665) x ABV74143 (1-1435)
 QY 240 ValLeuValHisCysLeuAlaGlyIleSerArgSer 251
 DB 1157 GTCTGTGTCACCTGCTGCAGGACATCACCCTCA 1192
 RESULT 41
 AADS7367
 ID AADS7367 standard; cDNA; 1471 BP.
 XX AADS7367;
 AC
 XX
 DT 06-NOV-2003 (first entry)
 XX
 DE Human kinase and phosphatase (KPP-40) cDNA.
 KW Human; kinase; phosphatase; KPP; cell proliferative disorder; hepatitis;
 KW atherosclerosis; cirrhosis; haemoglobinuria; polycythaemia vera; cancer;
 KW psoriasis; thrombocytopenia; developmental disorder; Reiter's syndrome;
 KW renal tubular acidosis; anemia; mental retardation; Alzheimer's disease;
 KW neurological disorder; Parkinson's disease; epilepsy; glomerulonephritis;
 KW autoimmune disorder; inflammatory disorder; AIDS; Goodpasture's syndrome;
 KW acquired immune deficiency syndrome; immunosuppressive; Crohn's disease;
 KW nodular; transgenic; dermatitis; multiple sclerosis; diabetes mellitus;
 KW allergy; gout; Grave's disease; Hashimoto's thyroiditis; bowel syndrome;
 KW osteoporosis; rheumatoid arthritis; Sjogren's syndrome; ophthalmological;

KW gene therapy; asthma; anticonvulsant; uropathic; pancreatitis; gene; ss.
 XX Homo sapiens.
 OS
 XX
 FH Key Location/Qualifiers
 FT CDS 170..690
 FT /*tag= a
 FT /product= "Human KPP"
 XX
 PN WO2003050084-A2.
 XX
 PD 19-JUN-2003.
 XX
 PF 06-DEC-2002; 2002WO-US039126.
 XX
 PR 07-DEC-2001; 2001US-0340235P.
 PR 19-DEC-2001; 2001US-0343007P.
 PR 21-DEC-2001; 2001US-0343546P.
 PR 04-FEB-2002; 2002US-0354388P.
 PR 15-FEB-2002; 2002US-0357675P.
 XX
 PA (INCY-1) INCYTE GENOMICS INC.
 XX
 PI Kabla AE, Chien D, Wilson AD, Swarnakar A, Gorvad AE,
 PI Hafalia AD, Emerling BM, Ramkumar J, Jin P, Griffin JA, Marquis JP,
 PI Baughn MR, Chawla NK, Leht-Mason PM, Khare R, Lee S, Hawkins PR,
 PI Becha SD, Lee SY, Sprague WM, Zebardjian Y;
 DR WPI: 2003-532894/50.
 DR P-PSDB; AAE37995.
 XX
 PS Claim 5; Page 281; 282pp; English.
 XX
 CC The invention relates to an isolated polypeptide, which is a human kinase
 CC and phosphatase (KPP). KPP agonists and antagonists are useful for
 CC diagnosing, treating or preventing disorders associated with aberrant
 CC expression of KPP, particularly cell proliferative disorders (e.g.
 CC arteriosclerosis, atherosclerosis, cirrhosis, hepatitis, paroxysmal
 CC nocturnal haemoglobinuria, polycythaemia vera, psoriasis, primary
 CC thrombocytopenia or cancer), developmental disorders (eg. renal tubular
 CC acidosis, anemia or mental retardation), neurological disorders (e.g.
 CC Alzheimer's disease, Parkinson's disease or epilepsy), autoimmune/
 CC inflammatory disorders (e.g. AIDS; acquired immune deficiency syndrome,
 CC allergies, asthma, autoimmune thyroiditis, contact dermatitis, Crohn's
 CC disease, diabetes mellitus, glomerulonephritis, Goodpasture's syndrome,
 CC gout, Grave's disease, Hashimoto's thyroiditis, irritable bowel syndrome,
 CC multiple sclerosis, osteoarthritis, osteoporosis, pancreatitis, Reiter's
 CC syndrome, rheumatoid arthritis, Sjogren's syndrome, uveitis), or viral,
 CC bacterial, fungal, parasitic, protozoan or helminthic infections. The KPP
 CC is useful in assessing the effects of exogenous compounds on the
 CC expression of nucleic acids and kinases and phosphatases. KPP gene is
 CC useful in gene therapy and for creating transgenic animals to model human
 CC disease. The present sequence is human KPP cDNA
 XX
 SQ Sequence 1471 BP; 295 A; 479 C; 384 G; 313 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 0.175 Length: 1471
 Score: 12.00 Matches: 12
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.80% Indels: 0
 DB: 8 Gaps: 0
 US-10-029-345A-109 (1-665) x AADS7367 (1-1471)
 QY 240 ValLeuValHisCysLeuAlaGlyIleSerArgSer 251
 ||||||||||||||||||||||||||||||||||||||||

XX New human kinases and phosphatases and polynucleotides, useful for
PT diagnosing, treating or preventing autoimmune or inflammatory disorders
PT (e.g. AIDS, allergy or anemia), multiple sclerosis, osteoarthritis,
PT cancer or hepatitis.
XX
PS Claim 5; SEQ ID NO 103; 424bp; English.
XX
CC The invention relates to a novel isolated polypeptide which is a human
CC kinase and phosphatase (KPP). The KPP polypeptides, polynucleotides,
CC agonists and antagonists are useful for diagnosing, treating or
CC preventing cell proliferative disorders such as atherosclerosis,
CC cirrhosis, hepatitis and cancer, developmental disorders e.g. mental
CC retardation, neurological disorders including Alzheimer's disease and
CC Parkinson's disease, autoimmune and inflammatory disorders such as
CC Crohn's disease and diabetes mellitus and finally, viral, bacterial,
CC fungal, parasitic, protozoan or helminthic infections. Furthermore, the
CC polynucleotides encoding KPP may be useful for creating transgenic
CC animals to model human disease, as well as during gene therapy
CC procedures. The current sequence is that of the human KPP cDNA of the
CC invention.
XX
SQ Sequence 1547 BP; 301 A; 507 C; 422 G; 317 T; 0 U; 0 Other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 0.184 Length: 1547
XX Score: 12.00 Matches: 12
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 1.80% Indels: 0
XX DB: Gaps: 0
XX
US-10-029-345A-109 (1-665) x ADC99150 (1-1547)
QY 240 ValLeuValHisCysLeuAlaGlyIleSerArgSer 251
DB 695 GTCTGTGTGACACTGCTGTGCAGGCATCAGCCGCTCA 730
XX
RESULT 44
AAD23608
ID AAD23608 standard; DNA; 1696 BP.
XX
AC AAD23608;
XX
DT 07-MAR-2002 (first entry)
XX
DE Human protein phosphatase-5 (PP-5) DNA.
XX
KW Human; protein phosphatase-5; PP-5; gene therapy; tranquilliser; amnesia;
KW arteriosclerosis; atherosclerosis; anxiety; anaemia; hepatitis; cataract;
KW amyotrophic lateral sclerosis; adenocarcinoma; cerebral palsy; psoriasis;
KW ulcerative colitis; myasthenia gravis; infection; schizophrenic disorder;
KW neurological disorder; epilepsy; neoplasm; Alzheimer's disease; dementia;
KW thyroiditis; dermatitis; diabetic mellitus; rheumatoid arthritis; stroke;
KW granulomatous disease; haemolytic anaemia; Crohn's disease; cancer; SCID;
KW severe combined immunodeficiency disease; immune system disorder; trauma;
KW developmental disorder; cell proliferative disorder; Addison's disease;
KW systemic lupus erythematosus; Parkinson's disease; myofibroblast; AIDS;
KW leukaemia; antiinflammatory; cirrhosis; muscular dystrophy; allergy;
KW chromosome 12; db.
XX
XX Homo sapiens.
XX
OS
XX
XX Key Location/Qualifiers
XX FH 19..1125
XX FT /*tag= a
XX FT /product= "Human protein phosphatase-5 (PP-5)"
XX
XX WO200181590-A2.
XX
XX 01-NOV-2001.
XX
XX 19-APR-2001; 2001WO-US012902.

XX
XX 20-APR-2000; 2000US-0199010P.
XX 05-MAY-2000; 2000US-0202340P.
XX 10-MAY-2000; 2000US-0203424P.
XX 18-MAY-2000; 2000US-0205642P.
XX 02-JUN-2000; 2000US-0208854P.
XX
XX (INCYTE GENOMICS INC.
XX
XX Tang YF, Yue H, Khan FA, Wang YE, Patterson C, Gandhi AR;
XX Walla NK, Stewart EA, Tribouley CM, Hafalia A, Nguyen DB;
XX Elliott VS, Lee EA;
XX
XX WPI: 2002-034445/04.
XX P-PSDB; AAE14240.
XX
XX Novel polypeptides for diagnosing, preventing, treating immune system,
XX neurological, developmental and cell proliferative disorders including
XX cancer, comprises protein phosphatase polypeptides and encoding
XX polynucleotides.
XX
XX Claim 5; Page 105; 105pp; English.
XX
XX
XX The invention relates to human protein phosphatases (PP-1 to PP-5) and
XX their corresponding DNA molecules. Protein phosphatases and their DNA's
XX are useful for diagnosis, treatment and prevention of immune system
XX disorders (AIDS, severe combined immunodeficiency disease (SCID), chronic
XX granulomatous disease, autoimmune haemolytic anaemia, Crohn's disease,
XX autoimmune thyroiditis, atopic dermatitis, diabetic mellitus, rheumatoid
XX arthritis, systemic lupus erythematosus, systemic sclerosis, Addison's
XX disease, ulcerative colitis, haemodialysis uremia, myasthenia gravis,
XX trauma, viral, bacterial, fungal, parasitic, protozoal and helminthic
XX infections); neurological disorders (epilepsy, stroke, cerebral neoplasm,
XX Alzheimer's disease, Huntington's disease, Parkinson's disease,
XX amyotrophic lateral sclerosis, dementia, prion diseases); developmental
XX disorders (Down syndrome, cerebral palsy, spinal cord diseases, amnesia,
XX autonomic nervous system disorders, muscular dystrophy, anxiety, anaemia,
XX schizophrenic disorders, diabetic neuropathy, Tourette's disorder,
XX cataract, sensorineural hearing loss); cell proliferative disorders
XX (burkitt's, arteriosclerosis, atherosclerosis, cirrhosis, hepatitis,
XX myelofibrosis, psoriasis, cancers including adenocarcinoma, leukaemia,
XX lymphoma, melanoma, myeloma, sarcoma and cancers of the bone, brain,
XX breast, ovary, bladder, heart, kidney, liver, and pancreas. Protein
XX phosphatases, its fragments and its antibodies are useful as elements on
XX a micro array which is useful to monitor or measure protein-protein
XX interactions, drug-target interactions and gene expression profiles. The
XX present sequence is human protein phosphatase-5 (PP-5) DNA which is
XX mapped to human chromosome 12
XX
SQ Sequence 1696 BP; 339 A; 560 C; 447 G; 350 T; 0 U; 0 Other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 0.201 Length: 1696
XX Score: 12.00 Matches: 12
XX Percent Similarity: 100.00% Conservative: 0
XX Best Local Similarity: 100.00% Mismatches: 0
XX Query Match: 1.80% Indels: 0
XX DB: Gaps: 0
XX
US-10-029-345A-109 (1-665) x AAD23608 (1-1696)
QY 240 ValLeuValHisCysLeuAlaGlyIleSerArgSer 251
DB 844 GTCTGTGTGACACTGCTGTGCAGGCATCAGCCGCTCA 879
XX
RESULT 45
AAC98059
ID AAC98059 standard; cDNA; 1848 BP.
XX
XX AAC98059;
XX
XX 09-MAR-2001 (first entry)
XX
XX

Query Match: 1.80% Indels: 0
 DB: 7 Gaps: 0
 US-10-029-345A-109 (1-665) x ABR2339 (1-2104)
 QY 240 ValLeuValHISCySLeuA1AG1Y1LeSeRArgSer 251
 DB 1225 GTCCTGTGATTCCTTGGCGGCATCAGCCGCTCC 1260

RESULT 47
 ADB53531
 ID ADB53531 standard; DNA; 2104 BP.
 XX
 AC ADB53531;
 XX
 DT 04-DEC-2003 (first entry)
 DE Primary rat hepatocyte toxicity modelling related gene SEQ ID NO:4073.
 XX
 KM toxic effect; gene expression profile; hepatotoxicity; diagnostic marker;
 KM toxicity marker; toxicity progression; drug screening;
 KM primary rat hepatocyte toxicity modelling; gene; ds.
 XX
 OS Rattus norvegicus.
 XX
 PN WO2003065993-A2.
 XX
 PD 14-AUG-2003.
 XX
 PF 04-FEB-2003; 2003WO-US003482.
 XX
 XX 04-FEB-2002; 2002US-0353171P.
 XX 13-MAR-2002; 2002US-036334P.
 XX 08-APR-2002; 2002US-0370248P.
 XX 10-APR-2002; 2002US-0371134P.
 XX 10-APR-2002; 2002US-0371135P.
 XX 11-APR-2002; 2002US-0371413P.
 XX 19-APR-2002; 2002US-0373601P.
 XX 19-APR-2002; 2002US-0373602P.
 XX 22-APR-2002; 2002US-0374139P.
 XX 08-MAY-2002; 2002US-0378370P.
 XX 09-MAY-2002; 2002US-0378652P.
 XX 09-MAY-2002; 2002US-0378653P.
 XX 09-MAY-2002; 2002US-0378665P.
 XX 09-JUL-2002; 2002US-0394230P.
 XX 09-JUL-2002; 2002US-0394253P.
 XX 04-SEP-2002; 2002US-0407688P.
 XX 28-JAN-2003; 2003US-0442900P.
 XX
 PA (GENE-) GENE LOGIC INC.
 XX
 PI Mendrick D, Porter M, Johnson K, Higgs B, Castle A, Orr M;
 PI Elshoff M;
 XX
 DR WPI; 2003-731472/69.
 XX
 PT Determining if a compound induces a toxic effect on a tissue or cell, for
 PT identifying hepatotoxic compounds, comprises comparing a gene expression
 PT profile of a tissue or cell sample to a database of tox mean and non-tox
 PT mean values.
 XX
 PS Claim 44; SEQ ID NO 4073; 874pp; English.
 XX
 CC The present invention describes a method for determining whether a
 CC compound induces a toxic effect on a tissue or cell. The method comprises
 CC preparing a gene expression profile of a tissue or cell sample exposed to
 CC the compound, and comparing the gene expression profile to a database
 CC comprising data or information on the tox mean and non-tox mean value.
 CC The method is useful for predicting or identifying at least one toxic
 CC effect, particularly hepatotoxicity, of a test or unknown compound. The
 CC genes listed in the specification are useful as diagnostic or toxicity
 CC markers for the prediction or identification of the physiological state

CC of tissue or cell sample that has been exposed to a compound, or to
 CC identify or predict the toxic effects of a compound or an agent. These
 CC may also be used as markers for monitoring toxicity progression or for
 CC drug screening. The present sequence represents a primary rat hepatocyte
 CC toxicity modelling related gene sequence from the present invention.
 XX
 SQ Sequence 2104 BP; 463 A; 559 C; 558 G; 524 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 0.247 Length: 2104
 Score: 12.00 Matches: 12
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 1.80% Indels: 0
 DB: 9 Gaps: 0
 US-10-029-345A-109 (1-665) x ADB53531 (1-2104)
 QY 240 ValLeuValHISCySLeuA1AG1Y1LeSeRArgSer 251
 DB 1225 GTCCTGTGATTCCTTGGCGGCATCAGCCGCTCC 1260

RESULT 48
 ACA56537
 ID ACA56537 standard; cDNA; 2109 BP.
 XX
 AC ACA56537;
 XX
 DT 06-JUN-2003 (first entry)
 DE Human signalling pathway polynucleotide probe SEQ ID NO 1135.
 XX
 KM Human; probe; ss; array element; Parkinson's disease;
 KM signalling pathway population; cancer; adenocarcinoma; leukaemia;
 KM immunopathy; AIDS; asthma; neuropathy; Alzheimer's disease; microarray.
 XX
 OS Homo sapiens.
 XX
 PN US6500938-B1.
 XX
 PD 31-DEC-2002.
 XX
 PF 30-JAN-1998; 98US-00016434.
 XX
 XX 30-JAN-1998; 98US-00016434.
 XX
 PR 30-JAN-1998; 98US-00016434.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Au-Young J, Sellhammer JJ;
 XX
 DR WPI; 2003-352189/33.
 XX
 PT Combination of polynucleotide probes, useful as array elements in a
 PT microarray for monitoring the expression of a number of target
 PT polynucleotides.
 XX
 PS Claim 1; SEQ ID NO 1135; 65pp; English.
 XX
 CC The invention relates to a combination which, comprises a number of
 CC polynucleotide probes comprising a sequence selected from one of the 1490
 CC sequences mentioned in the specification. The combination is useful as an
 CC array element in a microarray for monitoring the expression of a number
 CC of target polynucleotides. The microarray is particularly useful in the
 CC diagnosis and treatment of cancer and immunopathology and neuropathology.
 CC The microarray is useful in diagnostics and treatment regimens, drug
 CC discovery and development, toxicological and carcinogenicity studies,
 CC forensics and pharmacogenomics. The microarray is also useful for
 CC monitoring progression of diseases and for developing sophisticated
 CC profiles for the effects of currently available therapeutic drugs. The
 CC combination is also useful for purifying a subpopulation of mRNAs, cDNAs
 CC and genomic fragments and in research and diagnostic applications. The
 CC array can detect changes in expression in a large number of genes coding
 CC for different signalling pathway populations which can be used to diagnose

CC various diseases including cancer e.g. adenocarcinoma and leukaemia,
CC immunopathies e.g. AIDS and asthma, neuropathies e.g. Alzheimer's disease
CC and Parkinson's disease. The present sequence represents a polynucleotide
CC probe of the invention. Note: The sequence data for this patent did not
CC form part of the printed specification but was obtained in electronic
CC format directly from USPTO at
CC seqdata.uspto.gov/sequence.html?DocID=06500938B1
XX
SQ Sequence 2109 BP; 484 A; 555 C; 561 G; 509 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 0.248 Length: 2109
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: Gaps: 0

US-10-029-345A-109 (1-665) x ACAs6537 (1-2109)
QY 240 ValLeuValHisCysLeuAlaGlyIleSerArgSer 251
DB 1216 GCTTGATACATTGCTTGCTGCGATTAGCCGCTCA 1251

RESULT 49
ABK83457
ID ABK83457 standard; cDNA; 2390 BP.
XX
AC ABK83457;
XX
DT 14-AUG-2002 (first entry)
XX
DE Human cDNA differentially expressed in granulocytic cells #28.
XX
KW Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW viral infection; parasitic infection; protozoal infection;
KW fungal infection; sterile inflammatory disease; psoriasis;
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW adult respiratory distress syndrome; inflammatory bowel disease;
KW Crohn's disease; ulcerative colitis; periodontal disease;
KW granulocyte activation; chronic inflammation; allergy.
XX
OS Homo sapiens.
XX
FN WO200228999-A2.
XX
PD 11-APR-2002.
XX
PF 03-OCT-2001; 2001WO-US030821.
XX
PR 03-OCT-2000; 2000US-0237189P.
XX
PA (GENE-) GENE LOGIC INC.
XX
PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX
DR WPI; 2002-435328/46.
XX
PT Detecting granulocyte activation by detecting differential expression of
PT genes associated with granulocyte activation, which serves as diagnostic
PT markers that is useful for monitoring disease states and drug toxicity.
XX
PS Claim 1; SEQ ID NO 28; 114bp; English.
XX
XX The invention relates to detecting (M1) granulocyte (GC) activation
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC DNA chip analysis as given in the specification, and comparing the
CC expression level to an expression level in an unactivated GC, where
CC differential expression of Gs is indicative of GCA. Also included are
CC modulating (M2) GA by contacting GC with an agent that alters the
CC expression of at least one gene in Gs; (2) screening (M3) for an agent
CC capable of modulating GCA or an inflammation (especially chronic) in a

CC tissue, an allergic response in a subject, exposure of a subject to a
CC pathogen or sterile inflammatory disease using the gene expression
CC profile; (3) detecting (M4) an inflammation (especially chronic) in a
CC tissue, an allergic response in a subject, exposure of a subject to a
CC pathogen or sterile inflammatory disease, by detecting the level of
CC expression in a sample of the tissue of gene(s) from Gs, where the level
CC of expression of the gene is indicative of inflammation; (4) treating
CC (M5) an inflammation (especially chronic) or in a tissue, an allergic
CC response in a subject, exposure of a subject to a pathogen or sterile
CC inflammatory disease, by contacting a tissue having inflammation with an
CC agent that modulates the expression of gene(s) from Gs in the tissue. M1
CC is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful
CC for screening an agent capable of modulating GCA preferably in an
CC inflammation in a tissue; M4 is useful for detecting an inflammation
CC (especially chronic) in a tissue, an allergic response in a subject,
CC exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
CC psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, adult
CC cardiac reperfusion injury, renal reperfusion injury, ARDS, adult
CC respiratory distress syndrome, inflammatory bowel disease, Crohn's
CC disease, ulcerative colitis, periodontal disease, also bacterial
CC infection, viral infection, parasitic infection, protozoal infection,
CC fungal infection and M5 is useful for treating one of the above
CC conditions. The present sequence represents a gene differentially
CC expressed in granulocytes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 2390 BP; 565 A; 602 C; 616 G; 607 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 0.279 Length: 2390
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: Gaps: 0

US-10-029-345A-109 (1-665) x ABK83457 (1-2390)
QY 240 ValLeuValHisCysLeuAlaGlyIleSerArgSer 251
DB 1216 GCTTGATACATTGCTTGCTGCGATTAGCCGCTCA 1251

RESULT 50
ACC46774
ID ACC46774 standard; cDNA; 2390 BP.
XX
AC ACC46774;
XX
DT 05-JUN-2003 (first entry)
XX
DE Human COPD related protein encoding cDNA SEQ ID NO:25.
XX
KW Human; chronic obstructive pulmonary disease; COPD; chronic lung disease;
KW gene; ss.
XX
OS Homo sapiens.
XX
FN WO200297127-A2.
XX
PD 05-DEC-2002.
XX
PF 28-MAY-2002; 2002WO-EP005835.
XX
PR 31-MAY-2001; 2001GB-00013266.
XX
PA (FARB) BAYER AG.
XX
PI Oellers N, Gehrmann M, Kallabis H, Hall R, Schulze T, Kroegel C;
XX
DR WPI; 2003-140492/13.
XX
P-PSDB; ABP96803.

XX Predicting, diagnosing or prognosing chronic lung disease, by detecting a
PT chronic obstructive pulmonary disease (COPD) gene in a biological sample.
XX

PS Claim 8; Page 113-114; 214pp; English.

XX
CC The present invention describes a method for predicting, diagnosing or
CC prognosing chronic lung disease by detecting a chronic obstructive
CC pulmonary disease (COPD) gene related polynucleotide (see ACC46750 to
CC ACC46777, which encode the COPD related proteins in ABP96779 to
CC ABP96806). The method is useful for predicting, diagnosing or prognosing
CC chronic lung disease in a biological sample. The COPD genes and proteins
CC encoded by them from the present invention (I) can be used for treating
CC or preventing chronic lung disease in a mammal. (I) can be used in an
CC animal model for determining the efficacy, toxicity, or side effects of
CC treatment with (I), and determining the mechanism of action of (I).
CC ACC46778 to ACC46903 represent COPD related PCR primers and probes used
CC in an example from the present invention

XX SQ Sequence 2390 BP; 565 A; 602 C; 616 G; 607 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	0.279	Length:	2390
Score:	12.00	Matches:	12
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.80%	Indels:	0
DB:	7	Gaps:	0

US-10-029-345A-109 (1-665) x ACC46774 (1-2390)

QY	240	ValLeuValHisCysLeuAlaGlyIleSerArgSer	251
DB	1216	GTCCTGGTACATGCTTGCGCTGSCATTAGCGCGCTCA	1251

Search completed: June 22, 2004, 07:00:42
Job time : 764 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 22, 2004, 06:47:53 ; Search time 733 Seconds
(without alignments)
4156.104 Million cell updates/sec

Title: US-10-029-345A-109

Perfect score: 665

Sequence: 1 MAHEMGTGIVTERLVALL.....LKVSGSSFSGSMETIEVS 665

Scoring table:

OLIGO
Xgapop 60.0 , Xgapext 60.0
Ygapop 60.0 , Ygapext 60.0
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3017426 seqs, 2290544650 residues

Word size: 1

Total number of hits satisfying chosen parameters: 820462

Minimum DB seq length: 1419

Maximum DB seq length: 200000000

Post-processing: Listing first 75 summaries

Command line parameters:

-MODE=frame_plus_p2n.model -DEV=xlh
-Q/cgn2_1/USPTO.epool/US10029345/runat_21062004_164402_29793/app_query.fasta_1.839
-DB=Published Applications NA -QFWT=fastcap -SUFFIX=oligo.rmpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bites -START=1 -END=-1 -MATRIX=oligo
-TRANS=human40.cgi -LIST=75 -DOCCALIGN=200 -THR SCORE=quality -THR MIN=1
-ALIGN=50 -MODE=LOCAL -OUTWMT=ptc -NORM=ext -HEAPSIZES=500 -MINLEN=4419
-MAXLEN=200000000 -USER=US10029345_QCGN_1_1221@runat_21062004_164402_29793
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60
-FAPOP=6 -FAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Published Applications_NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq2:*
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
------------	-------	-------------	--------	----	-------------

1	472	71.0	1998	9	US-09-816-494-3	Sequence 27, Appl
2	472	71.0	1998	16	US-10-377-072-27	Sequence 673, App
3	472	71.0	2102	16	US-10-094-749-673	Sequence 2, Appl
4	472	71.0	2732	13	US-10-168-506-2	Sequence 520, App
5	472	71.0	2966	13	US-10-296-115-520	Sequence 1, Appl
6	472	71.0	3059	17	US-10-257-026-1	Sequence 20, Appl
7	472	71.0	3332	9	US-09-964-277-20	Sequence 1, Appl
8	472	71.0	3496	9	US-09-964-277-1	Sequence 1, Appl
9	472	71.0	3544	9	US-09-816-494-1	Sequence 25, Appl
10	472	71.0	3544	16	US-10-377-072-25	Sequence 26334, A
11	472	71.0	3625	13	US-10-425-114-26234	Sequence 17, Appl
12	472	71.0	3766	13	US-10-343-357-17	Sequence 115, App
13	472	71.0	4790	17	US-10-648-593-115	Sequence 255, App
14	408	61.4	2200	13	US-10-072-012-255	Sequence 257, App
15	304	45.7	2071	13	US-10-072-012-257	Sequence 2429, Ap
16	178	26.8	1916	16	US-10-108-260A-2429	Sequence 1, Appl
17	17	20	2453	14	US-10-005-858-1	Sequence 28, Appl
18	20	3.0	2476	13	US-10-220-120-28	Sequence 69, Appl
19	18	1.8	1848	9	US-09-925-299-69	Sequence 69, Appl
20	12	1.8	1848	10	US-09-925-299-69	Sequence 11, Appl
21	12	1.8	1852	16	US-10-199-221-11	Sequence 265, App
22	12	1.8	2039	13	US-10-072-012-265	Sequence 2041, Ap
23	12	1.8	2104	12	US-10-153-319A-2041	Sequence 654, App
24	12	1.8	2104	16	US-10-388-934-654	Sequence 270, App
25	12	1.8	2109	16	US-10-191-803-270	Sequence 1135, Ap
26	12	1.8	2109	16	US-10-305-720-1135	Sequence 946, App
27	12	1.8	2109	17	US-10-641-643-946	Sequence 698, App
28	12	1.8	2390	13	US-10-342-887-698	Sequence 698, App
29	12	1.8	2390	13	US-10-172-118-698	Sequence 183, App
30	12	1.8	2390	15	US-10-175-523-183	Sequence 4, Appl
31	12	1.8	2390	16	US-10-199-221-4	Sequence 4, Appl
32	12	1.8	2390	17	US-10-717-557-4	Sequence 48, Appl
33	12	1.8	2649	16	US-10-133-937-48	Sequence 48, Appl
34	12	1.8	2649	16	US-10-153-553-48	Sequence 1, Appl
35	12	1.8	2650	15	US-10-184-832-1	Sequence 69, Appl
36	12	1.8	2792	10	US-09-971-392-69	Sequence 330, App
37	12	1.8	2792	14	US-10-044-090-330	Sequence 12, Appl
38	12	1.8	7001	16	US-10-199-221-12	Sequence 241, App
39	10	1.5	16561	16	US-10-085-117-241	Sequence 423, App
40	40	1.4	1571	15	US-10-006-288-423	Sequence 801, App
41	9	1.4	1519	9	US-09-736-457-801	Sequence 801, App
42	9	1.4	1519	9	US-09-902-941-801	Sequence 801, App
43	9	1.4	1519	9	US-09-849-626-801	Sequence 801, App
44	9	1.4	1519	13	US-10-283-017-801	Sequence 801, App
45	9	1.4	1519	15	US-10-011-754-801	Sequence 801, App
46	9	1.4	1619	15	US-10-113-872-801	Sequence 51, Appl
47	9	1.4	1632	10	US-09-971-602-51	Sequence 1034, Ap
48	9	1.4	1685	14	US-10-002-600-51	Sequence 1034, Ap
49	9	1.4	1702	13	US-10-342-887-1034	Sequence 61, Appl
50	9	1.4	1702	13	US-10-172-118-1034	Sequence 60, Appl
51	9	1.4	1808	16	US-10-388-934-61	Sequence 31, Appl
52	9	1.4	1912	16	US-10-388-934-60	Sequence 648, App
53	9	1.4	1913	13	US-10-116-802-31	Sequence 276, App
54	9	1.4	1998	15	US-10-103-313-648	Sequence 276, App
55	9	1.4	2000	9	US-09-969-708-276	Sequence 10, Appl
56	9	1.4	2000	9	US-09-919-497-10	Sequence 3774, Ap
57	9	1.4	2000	9	US-09-880-107-3774	Sequence 1082, Ap
58	9	1.4	2000	10	US-09-960-706-1082	Sequence 1545, Ap
59	9	1.4	2000	13	US-10-240-425-1545	Sequence 825, App
60	9	1.4	2000	16	US-10-305-720-1291	Sequence 825, App
61	9	1.4	2015	16	US-10-341-434-92	Sequence 825, App
62	9	1.4	2064	9	US-09-736-457-825	Sequence 826, App
63	9	1.4	2064	9	US-09-849-626-825	Sequence 826, App
64	9	1.4	2064	9	US-09-902-941-825	Sequence 826, App
65	9	1.4	2064	13	US-10-283-017-825	Sequence 826, App
66	9	1.4	2064	15	US-10-017-754-825	Sequence 826, App
67	9	1.4	2064	15	US-10-113-872-825	Sequence 826, App
68	9	1.4	2109	9	US-09-736-457-826	Sequence 826, App
69	9	1.4	2109	9	US-09-902-941-826	Sequence 826, App
70	9	1.4	2109	9	US-09-849-626-826	Sequence 826, App
71	9	1.4	2109	13	US-10-283-017-826	Sequence 826, App
72	9	1.4	2109	15	US-10-017-754-826	Sequence 826, App

73	9	1.4	2109	15	US-10-113-872-826	Sequence 826, App
74	9	1.4	2170	9	US-09-925-302-204	Sequence 204, App
75	9	1.4	2170	13	US-09-925-302-204	Sequence 204, App

ALIGNMENTS

RESULT 1

```
US-09-816-494-3
; Sequence 3, Application US/09816494
; Patent No. US20020034807A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel A.
; TITLE OF INVENTION: 38692 AND 21117, NOVEL DUAL SPECIFICITY
; TITLE OF INVENTION: PHOSPHATASE MOLECULES AND USES THEREFOR
; FILE REFERENCE: 10448-030002
; CURRENT APPLICATION NUMBER: US/09/816,494
; PRIORITY FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 60/191,858
; PRIORITY FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1998
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-816-494-3
```

Alignment Scores:

Pred. No.:	0	Length:	1998
Score:	472.00	Matches:	663
Percent Similarity:	99.40%	Conservative:	0
Best Local Similarity:	99.40%	Mismatches:	2
Query Match:	70.98%	Indels:	4
DB:	9	Gaps:	0

US-10-029-345A-109 (1-665) x US-09-816-494-3 (1-1998)

```
QY 1 MetAlaHieGluMetIleGlyThrGlnIleValThGluArgLeuValAlaLeuEnglu 20
DB 1 ATGGCCCATGAGATGATTTGGAAGCTCAAAATTTTACTGAGAGGTTGGTGGCTCTGCTGGAA 60
QY 21 SerGlyThGluLysValLeuLeuIleAspSerArgProPheValGluTyraAsnThrSer 40
DB 61 AGTGGAAACGGAAGAAAGTGTGCTAAATTGATGACCGGCATTTTGGAAATACATCATCC 120
QY 41 HisIleLeuGluAlaIleAsnIleAsnIleAsnIleSerIleLeuMetIleArgArgLeuGln 60
DB 121 CACATTTTGGAGCCATTAATATCACTGCTCCAAAGCTTATGAAACGGAAGTTGCAACG 180
QY 61 AspLysValLeuIleThrGluLeuIleGlnIleSerAlaLysHisLysValAspIleAsp 80
DB 181 GACAAAGTGTATATTAACAGAGCTCAGCATTCAGCGAAACATTAAGGTTGACATTGAT 240
QY 81 CysSerGlnLysValValValIleAspGlnSerSerGlnAspValAlaSerLeuSerSer 100
DB 241 TCCAGTCAGAAAGTGTATGATTAACATCAAGCTCCCAAGATGTGGCTCTCTCTCTCA 300
QY 101 AspCysPheLeuThValLeuLeuGluLysLeuGluLysSerPheAspSerValHisLeu 120
DB 301 GACCTGTTTCTACCTGACTTCTGGGTAACTGGAGAAAGAGCTTCAACTCTGTTCACTG 360
QY 121 LeuAlaGlyGlyPheAlaGluPheSerArgCysPheProGlyLeuGluGlyLysSer 140
DB 361 CTTCGAGGTGGGTGTGCTGAGTCTCTGCTGTTTCCCTGGGCTCTGTGAGGAAGAAATCC 420
QY 141 ThrLeuValProThrCysIleSerGlnProCysLeuProValAlaAsnIleGlyProThr 160
DB 421 ACTCTAGTCCCTACCTGATTTCTCAGCCTTGCTTACCTGTTGCAACATTTGGGCAACC 480
QY 161 ArgIleLeuProAsnLeuTyrluGlyCysGlnArgAspValLeuAsnIleGluLeu 180
DB 481 CGAATTTCTCCCAATCTTATCTTGGCTGCGACGAGATGTCTTCACACAGGAGCTGAT- 539
```

```
QY 181 -GlnGlnAsnGlyIleGlyTyraValLeuAsnAlaSerTyra-ThrCysProLysProAsp 200
DB 540 GCACAGAAATGGGATGGTATGTGTAAATGCCAGCAA-TACCTGTCCAAAGCCTGACT 598
QY 200 HeIleProGluSerHisPheLeuArgValProValAsnAspSerPheCysGluLysIle 220
DB 599 TTATCCCCAGATCATTTCTCGCGTGGCCGTGTAAGACAGCTTTGTGAAATTTT 658
QY 220 euProTrpLeuAspLysSerValAspPheIleGluLysAlaLysAlaSerAsnGlyCys 240
DB 659 TCCCGTGGTGGACAAATCAGTAGATTCTTGAAGAAAGCAAAAGCCTCCATGATGTG 718
QY 240 AlLeuValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrlu 260
DB 719 TTCTAGTCCACTGTTTACCTGGGATCTCCCGCTCGCACCATGCTATCCCTACATCA 778
QY 260 eLysArgMetAspMetSerLeuAspGluAlaTyraArgPheValLysGluLysArgPro 280
DB 779 TGAAGAGGATGACATGCTTTAGATGAAGCTTACAGATTGTGAAAGAAAGAAAGACTTA 838
QY 280 hrIleSerProAsnPheAsnPheLeuGluGlnLeuLeuAspTyrgluLysLysIleLys 300
DB 839 CTATATCTCCAAACTTCAATTTTCTGGGCCCACTCTGGAATGAGAAAGATTTAAGA 898
QY 300 snGlnThrGlyAlaSerGlyProLysSerLysLysLeuLeuHisLeuGluLysPro 320
DB 899 ACCAGACTGGAGCATCAAGGGCCAAAGCAAACTCAAGCTGCTGACCTGGAGAAAGCCAA 958
QY 320 snGluProValProAlaValSerGluGlyGlnLysSerGluThProLeuSerPro 340
DB 959 ATGAACCTGCTCCCTGCTGTCTCAAGAGGTGACGAAAGGAGGAGCCCTCAGTCCAC 1018
QY 340 roCysAlaAspSerAlaThrSerGluAlaIleGlyGlnArgProValHisProAlaSer 360
DB 1019 CCTGTGCCGACTCTGTAACCTCAGAGGACAGAGCAAAAGGCCCTGTCATCCGCCAGCG 1078
QY 360 AlProSerValProSerValGlnProSerLeuLeuAspSerProLeuValGlnAla 380
DB 1079 TGCCAGAGGTGCCAGGCTGTGAGCCGCTGCTGTTAAGGAGAGCCCTGCTGACAGCGC 1138
QY 380 euSerGlyLeuHisLeuSerAlaAspArgLeuGluAspSerAsnLysLeuLysArgSer 400
DB 1139 TCAGTGGGCTGCACTGTCTGCGGACAGAGCTGGAAAGCAGCATTAAGCTTCAAGCTTCT 1198
QY 400 HeSerLeuAspIleLysSerValSerTyraSerAlaSerMetAlaAlaSerLeuHisGly 420
DB 1199 TCTCTGTGATATCAATCAGTTTCAATTCAGCCAGCATGGCAGCATCTTACATGGCT 1258
QY 420 HeSerSerSerGluAspAlaLeuGluTyrluTyrluProSerThrThrLeuAspGlyThra 440
DB 1259 TCTCTATCAGAAAGATCTTGGAAATCACTCAAACTTCCACTACTCTGGATGGAGCA 1318
QY 440 snLysLeuCysGlnPheSerProValGlnGluLeuSerGluGlnThrProGluThSerP 460
DB 1319 ACAAGTATGCCAGTTCTCCCTGTTCCAGGAACATTAAGAGAGACTCCGGAACAGATC 1378
QY 460 roAspLysGluGluAlaSerIleProLysLysLeuGlnThraAlaArgProSerAspSer 480
DB 1379 CTATATGAGGAGGAGGACAGCATCCCAAGAGCTGAGACCCGCAAGGCTTTCAGACAG 1438
QY 480 InSerLysArgLeuHisSerValArgThrSerSerSerGlyThrAlaGlnArgSerLeu 500
DB 1439 AGAGCAAGCATTTGCTTCCGTAGAAACAGGAGCTGACCGCCAGAGGATCTCCCTTT 1498
QY 500 euSerProLeuHisArgSerGlySerValGluAspAsnTyrluHisThrSerPheLeuPhe 520
DB 1499 TATCTCAGCTGATCGAAGGAGGAGCTGGAAGGACATTAACACACACACTTCTTTGG 1558
QY 520 LysLeuSerThrSerGlnGlnHisLeuThrluSerAlaGlyLeuGlyLysGlyTyrlu 540
DB 1559 GCTTTTCCACAGCCAGAGCAGCATTCACGAAGTGTGCTGCGCTTAAAGGGCTGAGC 1618
```

```

OY 5401aSeSaSpIleuMaIaPProGInThSeThProSeIeuthSeSerTPrTyPha 560
Db 1619 ACTCGaTATCTTGgCCCCCCCAAGACTCTAACCCCTTCCCTGACACAGCTGGATTTTG 1678
OY 5601aThGluSeSeSerhIaPheTySeSerIaSeRaIaIeTyGlyGlySeRaIaSeTyS 580
Db 1679 CCACAGAGCTCTACACTTCTACTCTGCTCTACGCCATCTACAGGAGCAGTCCAGTTACT 1738
OY 580 eRAlATySeTyCySeSeSerGInleuPProThCyGElYlApGInValTySeRaIaRyA 600
Db 1739 CTGCTTACAGCTGCGACCCAGCTGCTCCACTTGGCGAACAAGCTATTTCTGTGGCAGAGC 1798
OY 600 rGgInLySeProSeRaPaRaIaAPSeSeRaRgArSeSerTrpIaGInGluSeProPhG 620
Db 1799 GCGAGAGCCCAAGTACAGACTGACTCCCGCGGAGCTGCATGAAGAGAGCCCTTTTG 1858
OY 620 lUlyGInPheLySaRgArSeCyGElMeTcInPheGlyGluSerIlMeTSeRGlA 640
Db 1859 AAAACAGATTTAACCCAGCAGCTGCAATATGAAATTGGAGAGACATCATGTGCAGAGA 1918
OY 640 hARSeSerTrpGInGluSeuGlyLyVaGlySeRInSeSerPheSerGlySeMeTg 660
Db 1919 AAGGTCAAGGAGAGAGCTGGGGAAAGTGGGAGCTGATGATCTTTTGGGGCAGATGG 1978
OY 660 lUleIlleGluValSer 665
Db 1979 AAATCATTGAGGTCTCC 1995

RESULT 2
US-10-377-072-27
Sequence 27, Application US/10377072
Publication No. US20040009501A1
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals Inc.
APPLICANT: Curtis, Rory A.J.
APPLICANT: Logan, Thomas Joseph
APPLICANT: Glucksmann, Maria A.
APPLICANT: Meyers, Rachel E.
APPLICANT: Williamson, Mark J.
APPLICANT: Rudolph-Owen, Laura A.
APPLICANT: Chun, Miyoung
APPLICANT: Tsai, Fong-Ying
TITLE OF INVENTION: NOVEL 25869, 25934, 26335, 50365, 21117,
TITLE OF INVENTION: 36692, 46508, 16815, 16839, 49937, 49931 AND 49933 MOLECULES
FILE REFERENCE: MP103-018OWNIM
CURRENT APPLICATION NUMBER: US/10/377, 072
CURRENT FILING DATE: 2003-02-27
PRIOR APPLICATION NUMBER: US 09/895, 860
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: US 60/215, 370
PRIOR FILING DATE: 2000-06-29
PRIOR APPLICATION NUMBER: US 09/723, 806
PRIOR FILING DATE: 2000-11-28
PRIOR APPLICATION NUMBER: US 60/187, 455
PRIOR FILING DATE: 2000-03-07
PRIOR APPLICATION NUMBER: US 09/843, 297
PRIOR FILING DATE: 2001-04-25
PRIOR APPLICATION NUMBER: US 60/199, 801
PRIOR FILING DATE: 2000-04-26
PRIOR APPLICATION NUMBER: US 09/861, 801
PRIOR FILING DATE: 2001-05-21
PRIOR APPLICATION NUMBER: US 60/205, 508
PRIOR FILING DATE: 2000-05-19
PRIOR APPLICATION NUMBER: US 09/816, 494
PRIOR FILING DATE: 2001-03-23
PRIOR APPLICATION NUMBER: US 09/815, 419
PRIOR FILING DATE: 2001-03-22
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 114
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 27
LENGTH: 1998

```

	TYPE: DNA	ORGANISM: Homo Sapiens	FEATURE:	NAME/KEY: CDS	LOCATION: (1) ... (1998)	US-10-377-072-27
Alignment Scores:						
Pred. No.:	0	Length:	1998			
Score:	472.00	Matches:	663			
Percent Similarity:	99.40%	Conservative:	0			
Best Local Similarity:	99.40%	Mismatches:	2			
Query Match:	70.98%	Indels:	4			
DB:	16	Gaps:	0			
US-10-029-345A-109 (1-665) x US-10-377-072-27 (1-1998)						
QY	1	MetAlaHISgLUmeTlIeGlyThrGlnIleValThrGluArgPheValAlaLeuLeuGlu	20			
DB	1	ATGGCCCATGAGATGATTTGAAACTCAAAATTGTTACTGAGAGGTGGTGGCTGCTGGAA	60			
QY	21	SerGlyThrGlnIleValLeuLeuIleAspSerArgProPheValGluTyrAsnThrSer	40			
DB	61	AGTGAACGGAAGAAAGTGGCTCTTAATTGATTACCCGCCCTTTGTGAAATTCATATCATCC	120			
QY	41	HisIleLeuGluAlaIleAsnIleAsnCyserIleValMetIleValArgArgLeuGln	60			
DB	121	CACATTTTGAAGCCATTAAATATCACTGCTCCAACTTATGAAGCGAAGTTGCACAG	180			
QY	61	AspIleValIleIleThrGluLeuIleGlnHisSerAlaIleHisIleValAspIleAsp	80			
DB	181	GACAAAGTGTAAATTACAGAGCTCATCCAGCATTCAGCGAAACATTAAGTTGACATTGAT	240			
QY	81	CysSerGlnIleValIleValTyrAspGlnSerSerGlnAspValAsnIleSerIleSer	100			
DB	241	TGCAGTCAAGAGGTGTGATTACATCAAACTCCCAAGATGTGCTCTCTCTTCA	300			
QY	101	AspCysPheLeuThrValIleLeuGlnIleValSerGlnIleValSerPheIleSerValHisLeu	120			
DB	301	GACTGTTCCTCAGTACTCTTGAGTAACTGGAAGAAAGCTTCAACTGTTCACCTG	360			
QY	121	LeuAlaGlyIlePheAlaGluPheSerArgCysPheProGlyLeuGlyGlnIleIleSer	140			
DB	361	CTTGAGAGTGGGTTTGCTGAGTCTCTCTGTTTCCCTGCGCTCTGTGAAGAAATCC	420			
QY	141	ThrIleValProThrCysIleSerGlnProCysLeuProValAlaAsnIleGlyProThr	160			
DB	421	ACTCTAGTCCCTACCTGCACTTCTCAGCTTCTTCACTGTTGCCAAACATTTGGGCCCAACC	480			
QY	161	ArgIleLeuProAsnLeuTyrIleGlnIleCysGlnArgAspValLeuAsnIleGlnIle	180			
DB	481	CGAAATCTTCCCAATCTTATCTTGAGCTGCAGCGAAGTCTCTCAACAAAGAGCTGAT	540			
QY	181	GlnGlnAsnGlnIleGlyTyrValLeuAsnAlaSerTyrThrCysProIleProAsp	200			
DB	540	GCAGAGAAATGGGATTTGTTATCTGTAAATGCCAGCA-TACCTCTCCAAACCTGACT	598			
QY	200	HisIleProGlnSerHisPheLeuArgValProValAsnAspSerPheCysGlnIleIle	220			
DB	599	TTATCTCCGAGTCTATTTCTCGCTGCTGCTGTGAATACAGCTTTTGTGAAGAAATTT	658			
QY	220	EuProThrLeuAspIleSerValAspPheIleGlnIleValAlaValAsnAspGlyCysV	240			
DB	659	TGCCCTGGTGGCAAAATCAGTAGATTATAGAGAAACCAAAAGCTCCAAATGAGATG	718			
QY	240	AlaLeuValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIleM	266			
DB	719	TTCTATGTCACATGTTTACGTGGGATCTCCGCTCGCGCACACATCGCTATCCATCTCA	778			
QY	260	euIleValMetCaspMetSerIleAspGluAlaTyrArgPheValIleGlnIleValArgProT	280			
DB	779	TGAAGAGATGGAACATGCTTATAGTGAAGCTTACGATTTTGGAAAGAAAAGCTTA	838			

Db 296 TCGAGTCAGAGGTTGTAGTTTACGATCAAGAGTCCCAAGATGTTCCCTCTCTCTTCA 355
Qy 101 AAPPQVbPheLeuThrValLeuLeuGlyLybLeuGlnLysSerPheAsnSerValHisLeu 120
Db 356 GACTGTTTCTCACTGACTTCTGGGTAAACGTGGAAAGAGCTTCAACTGTTCTGACCTG 415
Qy 121 LeuAlaGlyGlyPheAlaGlyPheSerArgGlyPheProGlyLeuGlyGlyLysSer 140
Db 416 CTTCGAGGTGGGTTGCTAGATTCTCTGTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 475
Qy 141 ThrLeuValProThrCysIleSerGlnProCysLeuProValAlaAsnIleGlyProThr 160
Db 476 ACTCTAGTCCCTACCTGCACTTCTGACCTTGTTCACCTGTTCCCAACTGGGCCAACC 535
Qy 161 Arg-1LeuProAsnLeuTyriLeuGlyCysGlnArgAspValLeuAsnLysGlyLeu1 180
Db 536 CT-AATTTCTCCCAATCTTATCTTGTGGCTGGCCAGAGATGTCCTCAACAGAGCTGAT 594
Qy 180 e-GlnGlnaGnglyIleGlyTyrrValLeuAsnAlaSerTyrr-ThrCysProLysPProAsp 199
Db 595 -GCAGAGAGATGGATGGATTGTTAGTAAATGCCAGCAA-TACCTGTCCAAAGCCTGAC 652
Qy 200 PheIleProGlySerHisPheLeuArgValProValAsnAspSerPheCysGlyLysIle 219
Db 653 TTTATCCCGAGTCTCATTTCTGCGCTGTGCTGTAAATGACAGCTTTTGTGAGAAATTT 712
Qy 220 LeuProTrpLeuAspLysSerValAspPheIleGlyLysAlaLysAlaSerAnglyCys 239
Db 713 TTGCGGTGGTGGCAAAATCAGTAGATTTCATTGAGAAAGCAAAAGCTTCAATGATGAT 772
Qy 240 ValLeuValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrrIle 259
Db 773 GTTCTAGTCACCTGTTTACCTGGGATCTCCCGCTGCCACCACTGCTATCGCTACATC 832
Qy 260 MetLysArgMetAspMetSerLeuAspGlnAlaTyrrArgPheValLysGlyLysArgPro 279
Db 833 ATGGAAGAGATGCACTGCTTTAGATGAAGCTTACAGATTCTGAAAGAAAAGACCT 892
Qy 280 ThrIleSerProAsnPheAsnPheLeuGlyGlnLeuLeuAspTyrrGlyLysIleLys 299
Db 893 ACTATATCTCCAACTTCAATTTTCTGGGCAACTCTGGAATGAGAAGAAATTAAG 952
Qy 300 AsnGlnThrGlyAlaSerGlyProLysSerLysLeuLysLeuLeuHisLeuGlyLysPro 319
Db 953 AACCAAGACTGGAGCATCAGGCGCAAGCAAACTCAACTGCTGCACTGGAGAGCA 1012
Qy 320 AsnGluProValProAlaValSerGlyGlyGlnLysSerGlnThrProLeuSerPro 339
Db 1013 AATGAACCTGTCCTCTCTCTCTGAGAGGTGACAGAAAGCGAGCGCCCTCACTTCCA 1072
Qy 340 ProCysAlaAspSerAlaThrSerGlnAlaAlaGlyGlnArgProValHisProAlaSer 359
Db 1073 CCTGTGCGGACCTGCTCACTCAGAGGAGCAAGCAAAAGCGGTGATCCCGGCACAG 1132
Qy 360 ValProSerValProSerValGlnProSerLeuLeuGlnAspSerProLeuValGlnAla 379
Db 1133 GTGCCAGGCGTGGCCGCGTGCAGCGCTGCTTGAAGACAGCAATTAAGCTCAAGCGTCC 1192
Qy 380 LeuSerGlyLeuHisLeuSerAlaAspArgLeuGlnAspSerAsnLysLeuLysArgSer 399
Db 1193 CTCAGTGGGCTGCACTGTCCGACAGCGCTGGAAGACAGCAATTAAGCTCAAGCGTCC 1252
Qy 400 PheSerLeuAspIleLysSerValSerTyrrSerAlaSerMetAlaIleSerLeuHisGly 419
Db 1253 TTCTCTCTGGAATCAATCAGTTTCATATTCAGCCAGCATGGCAGCATCTTAACATGCG 1312
Qy 420 PheSerSerSerGlyAspAlaLeuGlyTyrrTyrrProSerThrThrLeuAspGlyThr 439
Db 1313 TTCTCTCTCAACAATGCTTTGGAACTACTCAAACTTCCACTCTGAGTGGAGACC 1372
Qy 440 AsnLysLeuCysGlnPheSerProValGlnLysSerGlnLysThrProGluThrSer 459

Db 1373 AACAGCTATGCCAGTCTCCCTGTTGAGAACTATCGAGACAGACTCCCAACCACT 1432
Qy 460 ProAspLysGlnGlnAlaSerIleProLysLysLeuGlnThrAlaArgProSerAspSer 479
Db 1433 CCGTATTAAGAGAAAGCCAGCATCCCAAGAGCTGCAATGCGGCTTTCAGACAC 1492
Qy 480 GlnSerLysArgLeuHisSerValArgThrSerSerSerGlyThrAlaGlnArgSerLeu 499
Db 1493 CAGAGCAAGCATTTGATTCGATCGATGACAGACAGACAGATGGACCGCCAGAGTCCCTT 1552
Qy 500 LeuSerProLeuHisArgSerGlySerValGlnAspAsnTyrrHisThrSerPheLeu 519
Db 1553 TTATCTCCACTGATCGAAGGTGGAGCGGTGAGGAAATTAACAACAACGATTCCTTTTC 1612
Qy 520 GlyLeuSerThrSerGlnGlnHisLeuThrLysSerAlaGlyLeuGlyLeuLysGlyTyrr 539
Db 1613 GGCCTTTCCACAGCAGCAGCAGCAGCCTCAAGAGTCTGCGCTGGCTTAAAGGCTGG 1672
Qy 540 HisSerAspIleLeuAlaProGlnThrSerThrProSerLeuThrSerSerTyrrPhe 559
Db 1673 CACTGGATATCTTGACCCCGCCAGACTTACCTTCCCTGACAGAGCTGGATTTT 1732
Qy 560 AlaThrGlySerSerHisPheTyrrSerAlaSerAlaIleTyrrGlyLysSerAlaSerTyrr 579
Db 1733 GCCACAGAGTCTCACTTCACTGCTGCTCAGGCATTAAGAGAGGAGTGCACATTAC 1792
Qy 580 SerAlaTyrrSerCysSerGlnLeuProThrCysGlyAspGlnValTyrrSerValArgArg 599
Db 1793 TCTGCTCAGCTGAGCGCCAGCTGCGCACTTCCGAGAACCAAGTCAATCTGTGCCACAG 1852
Qy 600 ArgGlnLysProSerAspArgAlaAspSerArgArgSerTyrrHisGlnGlySerProPhe 619
Db 1853 CGGCAAGAGCCAAAGTACAGACTGACTGCGGAGGTGGAGTGGAGTGAAGAGAGCCCTTT 1912
Qy 620 GlyLysGlnPheLysArgArgSerCysGlnMetGlnPheGlyGlySerIleMetSerGln 639
Db 1913 GAAAAGCATTTAAACGCAAGAGCTGCCAAATGGAATTTGAGAGAGCATGATGTACAG 1972
Qy 640 AsnArgSerArgGlnGlyLeuGlyLysValGlySerGlnSerSerPheSerGlySerMet 659
Db 1973 AACAGGTACCGGGAAGAGCTGGGGAAGTGGGAGTCACTTATGCTTTCGGGACAGATG 2032
Qy 660 GlnIleIleGlnValSer 665
Db 2033 GAATCATTTAGAGTCTCC 2050
RESULT 4
US-10-168-506-2
; Sequence 2, Application US/10168506
; Publication No. US20040053229A1
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY D.
; APPLICANT: MARTINEZ, RICARDO
; APPLICANT: WHYTE, DAVID
; APPLICANT: MANNING, GERARD
; APPLICANT: SUDARSANAM, SUCHA
; APPLICANT: HILL, RON
; APPLICANT: FLANAGAN, PETER
; TITLE OF INVENTION: MAMMALIAN PROTEIN PHOSPHATASES
; FILE REFERENCE: 038602/1351
; CURRENT APPLICATION NUMBER: US/10/168,506
; PRIOR FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: PCT/US00/34736
; NUMBER OF SEQ ID NOS: 76
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2732
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-168-506-2
Alignment Scores:

Pred. No.: 0
 Score: 472.00
 Percent Similarity: 99.40%
 Best Local Similarity: 99.40%
 Query Match: 70.98%
 DB: 13
 Length: 2732
 Matches: 663
 Conservative: 0
 Mismatches: 2
 Indels: 4
 Gaps: 0

US-10-029-345a-109 (1-665) x US-10-168-506-2 (1-2732)

QY 1 MetAlHisGluMetIleGlyThrGlnIleValThrGluArgLeuValAlaLeuLeuGlu 20
 DB 538 ATGGCCCATGAGATGATGATGACCTCAAAATGTTACTGAGAGGTGTGGCTGTGCTGGA 597
 QY 21 SerGlyThrGluHisValLeuLeuIleAspSerArgProPheValGluIlyrAntThrSer 40
 DB 598 AGTGAACCGAAGAAAGTGTGCTTAATGATGAGCCGACATTTGTGGAAATACATATCATCC 657
 QY 41 HisIleLeuGluValIleAsnIleAsnCySerIleLeuMetIlyArgArgLeuGlnGln 60
 DB 658 CACATTTTGGAGCCATTAATATCACTGCTCCAACTTATGAGGAAAGTTGGCAACAG 717
 QY 61 AspIlyValLeuIleThrGluLeuIleGlnHisSerAlaIlyHisIlyValAspIleAsp 80
 DB 718 GACAAAGTGTATTAACAGAGCTCATCCAGCAATTCAGCGAAACATAGGTTGACATTTGAT 777
 QY 81 CySerGlnIlyValIleValIlyrAspGlnSerSerGlnAspValAlaSerLeuSerSer 100
 DB 778 TGCAGTCAGAAAGTTGATTAACATACAAAGCTCCCAAGATGTTGCCCTCTCTCTTCA 837
 QY 101 AspCySerPheLeuThrValLeuLeuGlyIlyIleGluIlySerPheAsnSerValHisLeu 120
 DB 838 GACTGTTTCTCACTGACTCTGGGTAACTGGAGAAAGAGGTTCAACTCTGTCACTCG 897
 QY 121 LeuAlaGlyGlyPheAlaGluPheSerArgCySerPheProGlyLeuCyGluGlyIlySer 140
 DB 898 CTTGACGGGTGGTGGCTGAGTTCTCTCGTTGTTTCCCTGGCCCTGTGCAAGAAATCC 957
 QY 141 ThrLeuValProThrCysIleSerGlnProCysIleProValAlaAsnIleGlyProThr 160
 DB 958 ACTCTAGTCCCTACCTGCACTTTCTGACCTTGCTTACCTGTGGCCAACTTGGGCCAAC 1017
 QY 161 ArgIleLeuProAsnLeuThrLeuGlyCySerGlnArgAspValLeuAsnIlyGluLeu 180
 DB 1018 CCAATTTCTCCCAATCTTATCTTGGCTGCCAGCGAGATGCTCTCAACAGAGGCTGAT - 1076
 QY 181 -GlnGlnAsnGlyIleGlyTyrValLeuAsnAlaSerTyr-ThrCysProIlyProAsp 200
 DB 1077 GCAGCAGAAATGGATTTGTTATGTGTTAATGCCAGCA-TACCTGTCCAAAGCCTGACT 1135
 QY 200 HeIleProGluSerHisPheLeuArgValProValAsnAspSerPheCyGluIlyIle 220
 DB 1136 TTAATCCCGAGTCTCATTTCTCGCGGTGCCCTGTGAATACAGCTTTTGTGAAATAATT 1195
 QY 220 euProIlyLeuAspIlySerValAspPheIleGluIlyValIlyAlaSerAsnGlyCySer 240
 DB 1196 TGCCCTGGTGGCAAAATCAGTAGATTTCAATTGAGAAAGCAAAAGCCTCCATGATGAG 1255
 QY 240 AlLeuValHisCysIleLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIle 260
 DB 1256 TTTCTAGTGAATGTTAGCTGGGATCTCCGCTCCGCCACCAATCCCTATCCGCTTCATCA 1315
 QY 260 eLlyArgMetAspMetSerLeuAspGluAlaTyrArgPheValIlyGluIlyAspProT 280
 DB 1316 TGAAGAGATGACATGCTCTTTTNGATGAAGCTTACAGATTTGTGAAAGAAAGAAAGACCTA 1375
 QY 280 hriIleSerProAsnDheAsnPheLeuGlyGlnLeuLeuAspTyrGluIlyIlyIleIly 300
 DB 1376 CTATATCTCCAAACTTCAATTTTCTGGGCCAACTCCTGAGCTATGAGAAAGATTAAGA 1435
 QY 300 snGlnThrGlyAlaSerGlyProIlySerIlyLeuIlyLeuLeuHisIleGluIlyIlyPro 320
 DB 1436 ACCAGATGAGCATCAAGGCGCAAGAGCAAACTCAAGCTGCTGACCTGGAGAGGCCAA 1495

QY 320 snGlnProValProAlaValSerGluGlyGlnIlySerGluThrProLeuSerPro 340
 DB 1496 ATGAACCTGTCCCTCTCTCTCAAGAGGTGGACAGAAAGACAGACGCCCTCACTCAC 1555
 QY 340 roCyAlaAspSerAlaThrSerGluAlaAlaGlyIlyArgProValHisProAlaSer 360
 DB 1556 CCTGTGCCAGCTCTCTCACTCAAGAGGACAGAGCAAAAGGCCCTGTGATCCGCCAGCG 1615
 QY 360 AlProSerValProSerValGlnProSerLeuLeuGluAspSerProLeuValGlnAla 380
 DB 1616 TGCCAGAGGTGCCAGCTGGACGCCGTGTTTGAAGACAGCCCGCTGTTACAGGCGC 1675
 QY 380 euSerGlyLeuHisIleuSerIleAspArgLeuGluAspSerAsnIlyIlyIlyAspSer 400
 DB 1676 TCAGTGGCTGACCTGTCCGACAGGCTGGAGAGACAGCAATTAAGCTCAAGCGTTCT 1735
 QY 400 HeSerLeuAspIleIlySerValSerTyrSerAlaSerMetAlaIleSerLeuHisGlyP 420
 DB 1736 TCTCTGTGATATCAAACTAGTTTCAATTCAGCCAGCATGGCAGCATCTTACATGGCT 1795
 QY 420 HeSerSerSerGluAspAlaLeuGluTyrTyrIlyProSerThrThrLeuAspGlyThra 440
 DB 1796 TCTCTCATCAGAAAGATCTTTGGAATCTACAAACCTTCCACTACTCTGGATGGAGCA 1855
 QY 440 snIlyLeuCyGlnPheSerProValGlnGluIlySerGluIlyThrProGluThrSerP 460
 DB 1856 ACAGCTATGCGAGTTCTCCCTGTTCAGGAATATGAGAGACAGCTCCCAAGACAGATC 1915
 QY 460 roAspIlyGluGluAlaSerIleProIlyIlyIlyLeuGlnThrAlaArgProSerAspSer 480
 DB 1916 CTGATTAAGAGAGAGCCAGCATCCCAAGAAAGCTGCAAGCTCCAGAGCTTCAAGACACC 1975
 QY 480 InSerIlyArgLeuHisSerValArgThrSerSerSerGlyThrAlaGlnArgSerLeu 500
 DB 1976 AGAGCAAGCATGTGATTCGCTCAGAACCAAGACAGTGGACCGCCAGAGGTCCTTT 2035
 QY 500 euSerProLeuHisArgSerGlySerValGluAspAntTyrHisThrSerPheLeuPhe 520
 DB 2036 TATCTCCACTCATCGAATGAGAGCTGGAGGACAAATTAACACACAGCTTCTTTTCG 2095
 QY 520 IlySerSerThrSerGlnGlnHisIleuThrIlySerAlaGlyLeuGlyIlyLeuIlyIly 540
 DB 2096 GCTTTTCCACAGCAGCAGACGACCTCAAGAACTGTGGCTGGCTTAAAGGCTGGC 2155
 QY 540 iSerAspIleLeuAlaProGlnThrSerThrProSerLeuThrSerSerTyrPheAs 560
 DB 2156 ACTGGATATCTGGCCCCCAGACCTTACCCCTTCCCTGACAGAGCTGTATTTTG 2215
 QY 560 laThrGluSerSerHisPheTyrSerAlaSerAlaIleTyrGlyIlySerAlaSerTyr 580
 DB 2216 CCAGAGTCTTCACTTCTGCTTGTGCTTACGCACTTACAGAGGAGAGTCCAGTTACT 2275
 QY 580 eraIaTyrSerCysSerGlnLeuProThrCysGlyAspGlnValTyrSerValArgArg 600
 DB 2276 CTGCTTACAGCTGACAGCAGCTGCCCACTTGGCAGACCAAGCTTATCTGTGGCAGGC 2335
 QY 600 rGlnIlyProSerAspArgAlaAspSerArgAspSerTyrHisGlnGluIlySerProPhe 620
 DB 2336 GGCAGAAAGCCAAAGAGACAGCTGACTCGCGCGAGACTGGCAAGAGAGAGGCCCTTTG 2395
 QY 620 IlyIlyGlnPheIlyArgArgSerCysGlnMetGluPheGlyGluSerIleMetSerGlu 640
 DB 2396 AAAAGCAGTTTAAACGAGAAAGCTGCCAAATGAAATTTGAGAGACATCATGTCAAGAG 2455
 QY 640 snArgSerArgGluGluLeuGlyIlyValGlySerGlnSerSerPheSerGlySerMetG 660
 DB 2456 ACAAGTCACGGAGAGAGCTGGGAGAAAGTGGGCAAGTCACTAGCTTTCCGGGAGCATGG 2515
 QY 660 IulIleIleGluValSer 665
 DB 2516 AATCATTTGAGGTCTCC 2532

RESULT 5

	US-10-296-115-520	
/	Sequence 520, Application US/10296115	
/	Publication No. US20040053248A1	
/	GENERAL INFORMATION:	
/	APPLICANT: Hyseq Inc	
/	TITLE OF INVENTION: No. US20040053248A1el Nucleic Acids and Polypeptides	
/	FILE REFERENCE: 784PCT	
/	CURRENT APPLICATION NUMBER: US/10/296, 115	
/	CURRENT FILING DATE: 2002-11-18	
/	PRIOR APPLICATION NUMBER: US09/488, 725	
/	PRIOR FILING DATE: 2000-01-21	
/	PRIOR APPLICATION NUMBER: US09/552, 317	
/	PRIOR FILING DATE: 2000-04-25	
/	NUMBER OF SEQ ID NOS: 1478	
/	SEQ ID NO 520	
/	LENGTH: 2966	
/	TYPE: DNA	
/	ORGANISM: Homo sapiens	
/	FEATURE:	
/	NAME/KEY: misc_feature	
/	LOCATION: (1) ... (296)	
/	OTHER INFORMATION: n = a,t,c or g	
/	US-10-296-115-520	
	Alignment Scores:	
	Pred. No.: 0 Length: 2966	
	Score: 472.00 Matches: 648	
	Percent Similarity: 99.39% Conservative: 0	
	Best Local Similarity: 99.39% Mismatches: 2	
	Query Match: 70.98% Indels: 4	
	DB: 13 Gaps: 0	
	US-10-029-345A-109 (1-665) x US-10-296-115-520 (1-2966)	
OY	16 ValAlaLeuLeuGluSerGlyThrGluValLeuLeuLeuAspSerArgProPheVal	35
Dd	68 GTGGCTCTGCTGGAAAGTGAACGGAAAAGTGCTCTTAATTATAGCCGGCAATTTGTG	127
OY	36 GluTyraSnrThSerHisIleLeuGlnAlaIleAsnIleAsnCysSerIlysLeuMetLys	55
Dd	128 GAATACAATAKATCCACATTTTGGAAAGCATTAATACAACGTCCTCAAGCTTATAGAAG	187
OY	56 ArgAtgLeuGlnGlnAspLysValLeuIleThrGluLeuIleGlnHisSerAlaLysHis	75
Dd	188 CGAAGGTGCACACAGACCAAAGTGAATTACAGACTCATTACAGATTACAGCAAAAT	247
OY	76 LysValLeuAspIleAspCysSerGlnLysValValValTyraSpInSerSerGlnAspVal	95
Dd	248 AAGGTGACATTAATGATGACGTACAGAAAGTTGATGATTCATCAAGCTCCCAAGATGTT	307
OY	96 AlaSerLeuSerSerAspCysPheLeuThrValLeuLeuGluLysLeuGluLysSerPhe	115
Dd	308 GCCTCTCTCTCTTCACAGCTGTTTTCTCACCTGACTCTGGGTAAACTGGAGAAAGCTTC	367
OY	116 AsnSerValHisLeuLeuAlaIleGlyIlePheAlaGluPheSerArgCysPheProGlyLeu	135
Dd	368 AACCTCGTTCACCTGCTTCAGAGTGGGTTGAGAGTTCCTCGTTGTTCCCTGGGCTC	427
OY	136 CysGluGluLysSerThrLeuValProThrCysIleSerGlnProCysLeuProValAla	155
Dd	428 TGTGAAGGAATAATCCACTTAGTCCCTACCTGCAATTCCTACGCTTGCTTACCTGTTGCC	487
OY	156 AsnIleGluProThrArgIleLeuProAsnLeuTyrlleGluYCysGlnArgAspValLeu	175
Dd	488 AACATTGGGCGCAACCGAATTCCTCCCAATCTTATCTTGCTGGCTGCCAGCAATGTCCTC	547
OY	176 AsnLysGluLeuIle-GlnGlnAsnGlyIleGlyTyraValLeuAsnAlaSerTyr-ThrC	195
Dd	548 AACCAAGAGACTGAT-CGACGAGAAATGGGATGTTGTTATGCTTAAATGCCAGCAA-TACCT	605
OY	195 yepProLysProAspPheIleProGluSerHisPheLeuArgValProValAsnAspSerP	215
Dd	606 GTCCAAAGCTTGAATTCCTCCGAGATCTCAATTCCTGCGTGGCTGTGAATGACAGCT	665

QY	215	heCYsgLlyblyleleuProthleuapblysserValaapPhel1legluyval1aLySa	235
Db	666	TTTGTGAGAAATTTTGCGTGGTGGACAAATCAGTGAATTTCAATTGAGAAAGCAAAAG	725
QY	235	1aSerAenglyCYsValleuValh1scYseuValagly1leSer1Ser1aThrllea	255
Db	726	CCTCCAAATGATGTGTTCTAGTGACGTTAGCTGGGATCTCCGCTCGGCACCAATCG	785
QY	255	1a1le1a1rYr1leWellyaRgWetAphetSer1leuapglua1aTYrArghVal1	275
Db	786	CTATCGCCTCATCATGAGAGATGAGCAATGCTTTAAGAAAGCTTACAGATTTGGA	845
QY	275	YsgLlyblybArXProthrl1eSerProaPheanPheanPheleu1G1leu1euanpTYrg	295
Db	846	AAAGAAAAGACCTACTATATCTCAAACTCAATTTTCTGGGCCMACTCTGGAATATG	905
QY	295	1uLyblyb1eLybAengInThrlYal1aSerGlyProLybSerLybLeu1euanh	315
Db	906	AGAGAAGATTAAAGACCAAGCTGAGATCAGGACCAAGAAAGCAACTCAAGCTGTGC	965
QY	315	1sleuGlyblybProaenglybProVal1Proal1aVal1SerGlyGlyGlyblybSerGlyb	335
Db	966	ACCTGGAGAACCAATGAATCTGTCCCTGTCTCAGAGGGTGGACAGAAAGCGGA	1025
QY	335	hrProleuSerProProCYsAl1aapSer1aThrSerGlyAla1aGlyGlynaRgProX	355
Db	1026	CGCCCTCAGTCCACCTGTGGCCACTGTCTGCTACCTCAGAGGACGACGACAAAGCCCG	1085
QY	355	a1h1aPro1aSerVal1ProSerVal1ProSerVal1GlnProSer1leuenglybapSerP	375
Db	1086	TGCATCCCGCAGCGGTGCCAGCGTGCACGCGTGCAGCGCTGCTGTAGAGACAGACC	1145
QY	375	ro1euValGlnAl1aLeuSerGlyleu1h1sleuSer1a1aapRyLeuGlybapSer1eul	395
Db	1146	CGCTGGTACAGACCGCTCAGTGGCTGCACCTGTCCGACAGACGCTGGAAACGACATA	1205
QY	395	Ys1eulYbArXSerPheSer1leuapblyblybSerVal1SerTYSer1aSerWet1a1a	415
Db	1206	AGCTCAAGCCTCTCTCTCTGATTCAAATCAGTTTCATATTCACCGACGATGGCAG	1265
QY	415	1aSer1euanh1sGlyPheSer1SerGerglybapAl1aLeuGlyTYrlybProSerThrt	435
Db	1266	CATCCTTAACATGGCTTCTCCTCATCAGAAAGTCTTGGAATCAACAACTTCCACTA	1325
QY	435	hr1euanpRglybThra1nlyb1euanh1PheSerProVal1GlnGlyb1euanh1GlnT	455
Db	1326	CTCTGGATGGACCAACAAAGCTAAGCCAGTCTCCCTGTCAGAACTAATCGAGACGA	1385
QY	455	hrProGlybThrSerProapblybGlyGlnAl1aSer1leProLyb1euanh1Thrl1a1a	475
Db	1386	CTCCCGAAACCAAGTCTGATTAAGAGAGAGACGACATCCCAAGAGCTGACGACCGCA	1445
QY	475	rgProSerapSerGlnSerlybArgh1euanh1aSerVal1ArghThrSer1SerGlyThra	495
Db	1446	GGCCTTCAGACAGCAGACAGAGGATTTGATTGCTGACAGAACAGCAGCAGCTGGACCG	1505
QY	495	1aGln1aRgSer1leu1euanh1aRgSerGlyb1euanh1GlybapantYr1h1st	515
Db	1506	CCCGAGAGTCCCTTTATCTCCACTGATCAATGGAGAGGTGGAGAGACAAATTAACCA	1565
QY	515	hr1aRgPhe1euanh1euanh1SerGerglybGln1h1s1euanh1ThlybSer1aGly1euanh	535
Db	1566	CCAGCTTCCTTTTGCGCTTTCACCAAGCAGCAGCAGCAGCTCAAGATCTGCTGGCTGG	1625
QY	535	1y1euanh1yGlybThr1h1aSerapblyb1euanh1aProGlnThrSerThrProSer1euanh1st	555
Db	1626	GCCTTTAAGGGCTGGACATCGGATATCTTGCCCCCAAGACCTTAACCTTCCCTTGACCA	1685
QY	555	er1aRgPlybPhe1a1aThrGlybSer1h1aPheTYrSer1a1aSer1a1e1eYrGlyb	575
Db	1686	GCACCTGGATTTTGGCAGACAGAGTCTTCACCTTCTACTCTGCTCAGCCATCTAAGAG	1745

QY 480 InsertYsArGLeuHISerValArgThrSerSerSerGlyThrAlaGlnArgSerLeu 500
DB 1565 AGAGCAAGCATTCGATTCGGTGCAGAACGACGACGATGGCACCCGACAGATTCCTTT 1624
QY 500 euSerProLeuHISarGSerGlySerValGluAapAntYrHISerPheLeuPheG 520
DB 1625 TATCTCCATGCGATCGAAGTGGAGCGTGGAGGACATTCACACACAGCTTCTTTGG 1684
QY 520 LyLeuSerThrSerGlnGlnHISerThrLySerAlaGlyLeuGlyLeuGlyTyrH 540
DB 1685 GCGTTTCCACACACGACGACACACTCAGCAAGTCTGCTGGCGCTTAAAGCGCTGGC 1744
QY 540 ISerAapIleLeuAlaProGlnThrSerThrProSerLeuThrSerSerTyrPhe 560
DB 1745 ACTCGGATATCTGGGCCCGCCCAACCTTACCCCTTCCTGACACAGACGCTGATTTTG 1804
QY 560 IAThrGluSerSerHISpheTySerAlaSerAlaIleTyGlyGlySerAlaSerTyrS 580
DB 1805 CCAAGAGTCTCTACACTTCTACTGCTGCTCAGCCATCTACGAGGAGGAGTCCAGTTACT 1864
QY 580 exrIATyrSerCySerGlnLeuProThrCyGlyAapGlnValTyrSerValArgArg 600
DB 1865 CTGCTTACAGCTCAGCCAGCTCCCACTTGCAGAGACCAAGTCTATTCTGTCCGACAGC 1924
QY 600 rGgInLySProSerAapArgAlaAapSerArgArgSerTyrHISGluGluSerProPheG 620
DB 1925 GGCAGAAAGCCAAATGACAGAGTGACTCGCGCGAGAGCTGGCAATGAAGAGGCCCTTGG 1984
QY 620 IuLyGlnPheLySArGArGSerCyGlnMetGluPheGlyGluSerIleMetSerGlu 640
DB 1985 AAAAGCAGTTTAAACGAGAGAGCTGCCAAATGGAATTTGGAGAGATCATGTCAAGAG 2044
QY 640 smArGSerArgGluGluLeuGlyLyValGlySerGlnSerPheSerGlySerMetG 660
DB 2045 ACAGGTCACCGGAAAGCTGGGGAAGTGGGAGCAGTACGTACTTTCCGCGCAGCATGG 2104
QY 660 IuIleIleGluValSer 665
DB 2105 AAATCATTTAGAGTCTCC 2121
RESULT 7
US-09-964-277-20 : Sequence 20, Application US/09964277
Patent No. US20020137170A1
GENERAL INFORMATION:
APPLICANT: Luche, Ralf M.
ACT: We1, Bo
TITLE OF INVENTION: DSP-16 DUAL-SPECIFICITY PHOSPHATASE
FILE REFERENCE: 200125.434
CURRENT FILING DATE: 2001-09-25
CURRENT APPLICATION NUMBER: US/09/964,277
NUMBER OF SEQ ID NOS: 22
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 20
LENGTH: 3332.
TYPE: DNA
ORGANISM: Homo sapiens
US-09-964-277-20
Alignment Scores:
Pred. No.: 0 Length: 3332
Score: 472.00 Matches: 472
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 70.98% Indels: 0
DB: 9 Gaps: 0
US-10-029-345a-109 (1-665) x US-09-964-277-20 (1-3332)
QY 194 ThrCySProLySProAapPheIleProGluSerHISpheLeuArgValProValAaasp 213
DB 977 ACCTGTCAAAAGCTGATCTTATCCCGAGTCTCATTTCTGTGCTGTGTAATGAC 1036

QY 214 SerPheCyGluLyValLeuProTyrLeuAapLySerValAapPheIleGluLySAla 233
DB 1037 AGCTTTGTGAGAAATTTTGGCGTGGTGGACAAATCGATGATTTCAATGGAAAGCA 1096
QY 234 LySAlaSerAapGlyCyValLeuValHISCyLeuAlaGlyTyrLeSerArgSerAlaThr 253
DB 1097 AAAGCTCCAAATGAGATGTGTTTGTAGTCACTGTTTGTGGATGTCGCCGCTCCGACAC 1156
QY 254 IleAlaIleAlaTyrTlleMetLySArGMetAapMetSerLeuAapGluAlaTyrAapPhe 273
DB 1157 ATCGCATATGCTTACATCAATGAAGAGATGACATGTCTTTATATMACTTAACAATTT 1216
QY 274 ValLySGLuLyAapProThrIleSerProAapPheAapPheLeuGlyGlnLeuAap 293
DB 1217 GTGAAGAAAGAAAGACTCTATATCTCAAACTTAATTTTCTGGGCCAACTCTCGGAC 1276
QY 294 TyrGluLyLeuIleLySArGlnThrGlyAlaSerGlyProLySerLySLeuLySLeu 313
DB 1277 TATGAGAAAGATTTAAGAACCAAGCTGAGCATCAAGGCCAAAGCAAGCAAGCTG 1336
QY 314 LeuHISLeuGluLySProAangluProValProAlaValSerGluGlyGlnLySer 333
DB 1337 CTGCACTGAGAGAGCCAAATGACCTGTCCGTGCTCTCAGAGGGGTGACAGAAAGC 1396
QY 334 GluThrProLeuSerProProCySAlaAapSerAlaThrSerGluAlaIleGlyGlnArg 353
DB 1397 GAGAGCCCTCTAGTCACTGTGCGCACTGTGTACTCAGAGGACACAGCAAGCAAGG 1456
QY 354 ProValHISProAlaSerValProSerValProSerValGlnProSerLeuLeuGluAap 373
DB 1457 CCGTGATACCCGCGACGCTGCCACGCTGCCACGCTGACAGCGTCCGCTTTAGAGAGAC 1516
QY 374 SerProLeuValGlnAlaLeuSerGlyLeuHISLeuSerAlaAapArgLeuGluAapSer 393
DB 1517 AGCCCGCTGTACAGCGCTGAGTGGCTGCACCTGTCCGACAGACAGCTGGAAGACAGC 1576
QY 394 AamLySLeuLySArGSerPheSerLeuAapIleLySArGValSerTyrSerAlaSerMet 413
DB 1577 AAATAGCTCAAGGCTCTTCTCTGTGATTCAAATCAATTCATATTCAGCACACATG 1636
QY 414 AlaAlaSerLeuHISGlyPheSerSerSerGluAapAlaLeuGluTyrTyrLySProSer 433
DB 1637 GCAGATCTCTTACATGCTTCTCTCATCAGAAAGCTTTGGAAATCAACAACTTCC 1696
QY 434 ThrThrLeuAapGlyThrAamLySLeuCyGlnPheSerProValGlnLeuSerGlu 453
DB 1697 ACTACTCTGATGGAGCAACAAGCTATGCCAGTTCTCCCTGTTCAGGAACATATCGAG 1756
QY 454 GlnThrProGluThrSerProAapLySGLuAlaSerIleProLySLeuGlnThr 473
DB 1757 CAGACTCCGAAACCGATCTTAATAGAGAGAGCCAGATCCCAAGAGCTGCAAGACC 1816
QY 474 AlaArgProSerAapSerGlnSerLySArGLeuHISerValArgThrSerSerGly 493
DB 1817 GCCAGGCTTTCAGACGACCAAGCAAGCATTCGATTCGATCAAGAACGACGACATGGC 1876
QY 494 ThrAlaGlnArgSerLeuLeuSerProLeuHISarGSerGlySerValGluAapAntYr 513
DB 1877 ACCGCCCAAGGTCCTTTTATCTCCATCGATCGAAGTGGAGCGTGGAGGACAAATTAC 1936
QY 514 HisThrSerPheLeuPheGlyLeuSerThrSerGlnGlnHISerThrLySerAlaGly 533
DB 1937 CACACAGATTCCTTTTGGCGCTTTCACACACGACACACACTTCGAAAGCTGTGCTGGC 1996
QY 534 LeuGlyLeuLySGLyTyrHISerAapIleLeuAlaProGlnThrSerThrProSerLeu 553
DB 1997 CTGGGCTTAAAGGCTGGGACATCGGATATCTGGCCCCCAAGACTTACCTCTTCCCTG 2056
QY 554 ThrSerSerTyrTyrPheAlaThrGluSerSerHISpheTySerAlaSerAlaIleTy 573
DB 2057 ACCAGAGCTGTGATTTTCCACAGAGTCTCTACACTTCTACTCTGCTCAGGACATCTAC 2116

QY 574 G1GYSERAlaSerTyrSerAlaTyrSerCysSerGluLeuProThrCysGlyAspGln 593
DB 2117 GGAGGCAAGTGCAGTACTGCTGCTACAGCTGCAGCCAGCTGCCACTGGGAGACCAA 2176
QY 594 ValTyrSerValArgArgArgGlnLysProSerAspArgAlaAspSerArgSerTyr 613
DB 2177 GTCATTTCTGTGCGCAGCGCGCAGAAAGCCAGAGTGCAGAGCTGACTGCCGCGGAGCTGG 2236
QY 614 H1SG1GUSerProPheGluLysGlnPheLysArgArgSerCysGlnMetGluPheGly 633
DB 2237 CATGAGAGAGCCCTTTGAAAGACAGTTTAAAGCAGAGAGCTGCCAAATGGAAATTTGA 2296
QY 634 GluSer11MetSerGluAsnArgSerArgGluGluLeuGlyLysValGlySerGlnSer 653
DB 2297 GAGAGCATCATGTCAAGAGAACAGTCAACGGAAAGCTGGGAGAAAGTGGCAGTCACTT 2356
QY 654 SerPheSerGlySerMetGluLysLeuGluValSer 665
DB 2357 AGCTTTTCGGCAGCATGAAATCATTTAGAGTCTCC 2392

RESULT 8

US-09-964-277-1
: Sequence 1, Application US/09964277
: Patent No. US20020137170A1
: GENERAL INFORMATION:
: APPLICANT: Lucbe, Ralf M.
: APPLICANT: Wei, Bo
: TITLE OF INVENTION: DSP-16 DUAL-SPECIFICITY PHOSPHATASE
: FILE REFERENCE: 200125 434
: CURRENT APPLICATION NUMBER: US/09/964, 277
: NUMBER OF SEQ ID NOS: 22
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 1
: LENGTH: 3496
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-964-277-1

Alignment Scores:
Pred. No.: 0 Length: 3496
Score: 472.00 Matches: 663
Percent Similarity: 99.40% Conservative: 0
Best Local Similarity: 99.40% Mismatches: 2
Query Match: 70.98% Indels: 4
DB: 9 Gaps: 0

US-10-029-345A-109 (1-665) x US-09-964-277-1 (1-3496)

QY 1 MetAlaHisGluMetLeuGlyThrGlnLeuValThrGluArgLeuValAlaLeuLeuGlu 20
DB 562 ATGGCCCAATGAGATGATGAACTCAAAATGTTACTGAGAGGTTGCTGCTCTGAGAA 621
QY 21 SerGlyThrGluLysValLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 40
DB 622 AGTGAAGACGAAAGAAAGTCTGCTCAATGATGATACCCGACCAATTTGGAAATACAAATCATCC 681
QY 41 HisLeuLeuGluAlaIleAsnIleAsnCysSerLysLeuMetLysArgArgLeuGlnGln 60
DB 682 CACATTTTGGAGCATTATATCAATCACTGCTCCAAAGCTTATGAAAGCGAAAGTTGCAACG 741
QY 61 AspLysValLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 80
DB 742 GACAAAGTGTATTAACAGAGCTCATCCAGCAATTCAGGAAACATTAAGTTGACATTGAT 801
QY 81 CysSerGlnLysValValValValTyrAspGlnSerSerGlnAspValAlaSerLeuSerSer 100
DB 802 TGCACATCAGAGAGGTGTGTTTACATTAAGAGCTCCAGAGTGTGCTCTCTCTTCA 861
QY 101 AspCysPheLeuThrValLeuLeuLeuLysLysLeuGluLysSerPheAsnSerValHisLeu 120
DB 862 GACTGTTTCTCACTGTAATCTTCTGGGTAAACTGAGAAAGAGCTTCACTCTGTTCACTG 921

QY 121 LeuAlaGlyGlyPheAlaGluPheSerArgCysPheProGlyLysCysGluGlyLysSer 140
DB 922 CTGGAGAGTGGGTTTGCTGATAGTCTCTGCTGTTCTTCCCTGGCTCTGTGAAAGAAATCC 981
QY 141 ThrLeuValProThrCys11LeuSerGlnProCysLeuProValAlaAsnIleGlyProThr 160
DB 982 ACTTAATGCTTCACTGATCTTCAAGCTTGTCTTACCTGTTGCTTCAATATGGGCAACC 1041
QY 161 ArgIleLeuProAsnLeuTyrLeuGlyCysGlnArgAspValLeuAsnLysGluLeuIle 180
DB 1042 CGAATTTCTTCCAACTTATTTCTTGGCTGCGCAGAGATGTTCTCAACAGAGAGCTGAT- 1100
QY 181 -GlnGlnAsnGly11LeuGlyValLeuAsnAlaSerTyr-ThrCysProLysProAsp 200
DB 1101 GCAGCAGATGGATGGATTATGTTTAATGCCAGCAA-TACCTGTCAAAAGCTGACT 1159
QY 200 He11ProGluSerHisPheLeuArgValProValAsnAspSerPheCysGluLysIleL 220
DB 1160 TTATCCCGAGTCTATTTCTGCGTGTGCTGTGAATGACAGCTTTGTGAGAAATTT 1219
QY 220 euProTyrLeuAspLysSerValAspPhe11LeuLysAlaLysAlaSerAsnGlyCysV 240
DB 1220 TGCCTGTGTGGCAAACTACATGATTTTCAATTGAGAAAGCAAAAGCTTCCATGATGATG 1279
QY 240 AlLeuValHisCysLeuAlaGlyLysSerArgSerAlaThrIleAlaIleAlaTyrIleM 260
DB 1280 TTCTAGTGCACCTGTTTGTGCTGGATCTCCGCTCCGACATGCTATGCTCATCATCA 1339
QY 260 eLysArgMetAspMetSerLeuAsnGluAlaTyrArgPheValLysGluLysLeuArgProT 280
DB 1340 TGAAGAGATGAGATGCTTCTTATGATGAGAGCTTACATTTGTGAAGAAAGAAAGACTTA 1399
QY 280 hr11SerProAsnPheAsnPheLeuGlyGlnLeuLeuAspTyrGluLysIleLysA 300
DB 1400 CTATATCTCCAACTTCAATTTCTGCGCCAACTCTCGACTATGAGAAAGATTTAGA 1459
QY 300 snGlnThrGlyAlaSerGlyProLysSerLysLeuLysLeuHisLeuGluLysProA 320
DB 1460 ACCAGACTGAGAGATCAAGGCGCAAGAGCAAACTCAAGCTGTCACCTGGAAGAGCCAA 1519
QY 320 snGluProValProAlaValSerGluGlyGlyGlnLysSerGluTyrProLysSerProp 340
DB 1520 ATGAACCTGTCCTCGCTGCTCAGAGGGGTGAGAGAAAGCGAGCCCTTCAATGCCAC 1579
QY 340 rocValAlaAspSerAlaThrSerGluAlaAlaGlyGlnArgProValHisProAlaSerV 360
DB 1580 CCGTGTCCGACTCTGCTACCTCAGAGGAGAGCAAGGAGCCGTCATCCGCGCAGCG 1639
QY 360 alProSerValProSerValGlnProSerLeuLeuGluAspSerProLeuValGlnAla 380
DB 1640 TGCCCAAGTGGCCCAAGCTGACCGCTGCTGTGTAAGAGACAGCCGCTGGATACAGGCTC 1699
QY 380 euSerGlyLeuHisLeuSerAlaAspArgLeuGluLysSerAsnLysLeuLysArgSerP 400
DB 1700 TCAGTGGCTGACACTGTCGAGAGCAGGCTGGAAACAGCAATTAAGCTCAAGCTTCT 1759
QY 400 heSerLeuAspIleLysSerValSerTyrSerAlaSerMetAlaAspLeuHisGlyP 420
DB 1760 TCTCTCTGATATCAATCAATGATTTATCTTACGCCAGCATGCGAGATCTTATCAATGCT 1819
QY 420 heSerSerSerGluAspAlaLeuGluTyrTyrLysProSerThrThrLeuAspGlyLysP 440
DB 1820 TCTCTCTCAACAGAGATGCTTGGAACTTCAAACTTCCACTCTCTGAGTGGAGCA 1879
QY 440 snLysLeuCysGlnPheSerProValGlnGluLeuSerGluGlnThrProGluThrSerP 460
DB 1880 ACAAGCTATGCAAGTCTTCCCTGTTCAAGAACTATCGAGAGCACTCCGAAACCAAGTC 1939
QY 460 roAspLysGluGluAlaSerIleProLysLysLeuGlnThrAlaArgProSerAspSerG 480
DB 1940 CTGATTAAGAGAGAGCAATCCCAAGAGCTGCAAGCCGCGGCTTCAAGCAACC 1999
QY 480 InSerLysArgLeuHisSerValArgThrSerSerSerGlyThrAlaGlnArgSerLeuL 500

```

DB 2000 AGGAGAGGATTTGATTCGGTTCAGAAACGACGACGATGGSCACCGCCACGAGGTCCTTT 2059
QY 500 euSerProLeuHi.ea.rgSerGlySerValGluAapAntyRhi.eThrSerPheLeuPheG 520
DB 2060 TATCTCCATCGCATCCAAAGTGGAGCGTGGAGACATTAACACACAGCTTCTTTGCG 2119
QY 520 lYeuSerThrSerGlnGlnHisLeuThrlYSerIaGlyLeuGlyLeuValGlyTyrH 540
DB 2120 GCTTTTCACACACGACGACACCTCAAGAGTCTGCTGCGCTTAAAGGCTGGC 2179
QY 540 iSerAapIleLeuAlaProGlnThrSerThrProSerLeuThrSerSerIleTyrPheA 560
DB 2180 ACTGGATATTCCTGGCCCCCACCACCTTACCTTCCCTGACACGACGATGTTATTTG 2239
QY 560 lAthrGluSerSerHi.ePheTyrSerAlaSerAlaIleTyrGlyGlySerAlaSerTyrS 580
DB 2240 CCAAGAGTCTCAGACTTCTACTGCTTGCCTCAGCCATCTACGAGGACAGTCCAGTTACT 2299
QY 580 ePrlAthrSerCySerSerGlnLeuProThrCyGlyIleAapGlnValTyrSerValArgArgA 600
DB 2300 CTGCTTACAGCTGACACGACGCTCCACTTGGGAGACCAAGTCTATCTGTGCGAGCG 2359
QY 600 rGlnIlySProSerAapArgAlaAapSerArgArgSerTyrPheIaGluGluSerProPheG 620
DB 2360 GCGAAGACCAACTGACAGAGCTGACTCGCGGCGGAGCTGACATGAAGAGCCCTTTG 2419
QY 620 lYleuGlnPheIyA.rg.rgSerCySerGlnMetGluPheGlyGluSerIleMetSerGluA 640
DB 2420 AAAAGCAGTTTAAACGACAGAGCTGCCAAATGGAAATTGGAGAGACATCATGTGACAGAG 2479
QY 640 aP.rgSer.rg.rgIleuGlyLeuValGlySerGlnSerSerPheSerGlySerMetG 660
DB 2480 ACAGGTCACGGAAGAGCTGGGAAAGTGGGACGTACGTACTTTTCGGGACAGATG 2539
QY 660 lYleIleGluValSer 665
DB 2540 AATCATTTGAGCTCTCC 2556

```

RESULT 9

```

US-09-816-494-1
; Sequence 1, Application US/09816494
; Patent No. US2002034807A1
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel A.
; TITLE OF INVENTION: 38692 AND 21117, NOVEL DUAL SPECIFICITY
; FILE OF INVENTION: PHOSPHATASE MOLECULES AND USES THEREFOR
; FILE REFERENCE: 10448-030002
; CURRENT APPLICATION NUMBER: US/09/816,494
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 60/191,858
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3544
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (589)...(12583)
US-09-816-494-1

```

Alignment Scores:

```

Pred. No.: 0 Length: 3544
Score: 472.00 Matches: 663
Percent Similarity: 99.40% Conservative: 2
Best Local Similarity: 99.40% Mismatches: 0
Query Match: 70.98% Indels: 4
DB: 9 Gaps: 0

```

US-10-029-345A-109 (1-665) x US-09-816-494-1 (1-3544)

```

QY 1 MetAlaHisGluMetIleGlyThrGlnIleValIThrGluArgLeuValAlaLeuLeuGlu 20
DB 589 ATGGCCCATGATGATGATTTGAACTCAAAATTTTACTGAAAGGTTGGCTCTGCTGGAA 648
QY 21 SerGlyThrGluValLeuLeuIleAapSerArgProPheValGluTyrAanthrSer 40
DB 649 AGTGAAGAGGAAAAAGTGTGCTGAATGATGATGCCGCGCAATTTGTGAATACATATACCC 708
QY 41 HisIleLeuGlnAlaIleAanIleAanCySerIyLeuMetIyA.rg.rgLeuGlnGln 60
DB 709 CACATTTTGAAGCCATTAATATCACTCTCCAGCTTAATGAAAGGTTGCAACAG 768
QY 61 AapIyValLeuIleThrGluLeuIleGlnHisSerAlaIleHisIyValAapIleAap 80
DB 769 GAACAAGTGTAAATACAGAGCTCATCCAGCATTCAGCAACATAAAGTTGACATTTGAT 828
QY 81 CySerGlnIyValValValTyrAapGlnSerSerGlnAapValAlaSerLeuSerSer 100
DB 829 TCGAGTCAGAGGTTGTATTTACATCAAGTCCCAAGATGTTCCCTCTCTTCA 888
QY 101 AapCyPheLeuThrValLeuLeuGlyIyLeuGlnIySerPheAanSerValHisLeu 120
DB 889 GACTGTTTTCTCAGTGTACTTCTGGGTTAAACTGAGAAAGGCTTCAACTCTGTTCACTG 948
QY 121 LeuAlaGlyIyPheAlaGluPheSerArgCyPheProGlyLeuCyGluGlyIySer 140
DB 949 CTTCAGAGTGGGTTGTGATGTTCTCTGTTTCTTCCGCTCTGTGAAGGAAATATCC 1008
QY 141 ThrLeuValProThrCyHisIleSerGlnProCyLeuProValAlaAanIleGlyProThr 160
DB 1009 ACTTAGTCCCTACCTGACATTTCTGACCTTCTTACCTGTTGCCAACATTTGGGCCAAC 1068
QY 161 ArgIleLeuProAanLeuTyrLeuGlyCyGlnA.rgAapValLeuAanIyGluLeu 180
DB 1069 CGAATTTCTCCAACTTTATCTGCTGCTCCAGCGAGTGTCTCAACAGAGGCTGAT- 1127
QY 181 -GlnGlnAanGlyIleGlyTyrValLeuAanIaSerTyr-ThrCySPolyAProAap 200
DB 1128 GCGACGAATGGATGTTATGTGTTAATGCCAGCAA-TACCTGTCCAAAGCCGTGACT 1186
QY 200 HeIleProGluSerHisPheLeuArgValProValAanAapSerPheCyGluIyIle 220
DB 1187 TTTATCCCGCAGTCTATTTCTCGCGTGTCTGTGAATACAGCTTTTGTGAAAAATTT 1246
QY 220 euProTyrLeuAapIySerValAapPheIleGluIyAlaIyAlaSerAanGlyCyA 240
DB 1247 TCCCGTGGTTGACAAATCAGTAGATTTCAATGAGAAACAAAGCCCTCAATGATGTC 1306
QY 240 AlLeuValHisCySerLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIle 260
DB 1307 TTTCTAGTGCACGTGTTAGCTGGGATCTCCGCTCCGCCACCAATCGCTATCGCTATAC 1366
QY 260 eTlyA.rgMetAapMetSerLeuAapGluAlaTyrA.rgPheValIyGlyIyA.rgPro 280
DB 1367 TGAAGAGATGACATGCTTTTAGATGAAGCTTACAGATTTGTGAAGAAAAAGACCTTA 1426
QY 280 hrIleSerProAanPheAanPheLeuGlyGlnLeuAanApyTyrGluIyIyValIyA 300
DB 1427 CTATATCTCAAACTTCAATTTTTCGGGCAACTCCTGAGATATGAGAGAAATTAAGA 1486
QY 300 aGlnIThrGlyAlaSerGlyProIySerIyLeuLeuIleLeuHisIleLeuGlyIyPPro 320
DB 1487 ACCAGACTGGACATCAGGCGCAAGAGCAAACTCAAGTGTGTGACCTGAGAGAACCA 1546
QY 320 aGluProValProAlaValSerGluGlyIyGlyIyIySerGluThrProLeuSerPro 340
DB 1547 ATGAACCTGTCTCTGTCTGTAGAGGTTGACAAAGAACAGAGCCCTCTCATGTCAC 1606
QY 340 roCyAlaAapSerAlaThrSerGluAlaIleGlyGlnA.rgProValHisProAlaSerV 360
DB 1607 CCTGTGCGGACTCTGACTCAGAGGCGAGCAAAAGCCCGTGCATCCGCGCAGCG 1666
QY 360 aIProSerValProSerValGlnProSerLeuLeuGluAapSerProLeuValGlnAla 380

```

```

Db      1667 TGCCAGCGTCCCGCCGCGTGCAGCGCTGCTTGAAGAGACAGCCCGCTGTACAGCCG 1726
Qy      380 euserGlyLeuHiSleuserAlaAspArgLeuGluAspSerAsnLysLeuLysArgSerp 400
Db      1727 TCAGGGGCTCACCCTGTCCGACAGACAGCTGGAAAGCAGCAATTAAGCTCAAGGCTTCT 1786
Qy      400 heserLeuAspIleLysSerValSerYrSerAlaSerMetAlaAlaSerLeuHiSglYp 420
Db      1787 TCTCTGCGATATCAATCACTTTCATATTCAGCCAGATGAGCAATCTTACATGCT 1846
Qy      420 heserSerSerGluAspAlaLeuGluYrYrLysProSerThrThrLeuAspGlyThra 440
Db      1847 TCTCTCATCAGAAAGATGCTTTGGAAATACAAACCTTCACACTCTGGATGGAGCCA 1906
Qy      440 snLysLeuCYsGlnPhSerSerProValGlnGluLeuSerGluGlnThrProGlnThrSerp 460
Db      1907 ACAAGCTATGCCAATTCCTCCCTGTTGAGAAACTATGGAGCAGACTCCGAAACCAAGTC 1966
Qy      460 roAspLysGluGluAlaSerIleProLysLysLeuGlnThrAlaArgProSerAspSerG 480
Db      1967 CTGATTAAGAGAAAGCCAGCATCCCGAAGAGCTGCAGACCGCCAGGCTTCAGACAGCC 2026
Qy      480 InserLysArgLeuHiServalArgThrSerSerSerGlyThrAlaGlnArgSerLeu 500
Db      2027 AGAGCAAGCAGATTCATTCGCTCAGAACCAAGCAGAGTGGCAGCCCGCCAGAGGTCCTTT 2086
Qy      500 euserProLeuHiSArgSerGlySerValGluAspAsnYrHiSThrSerPheLeuPhag 520
Db      2087 TATCTCCACTGCATCAGAGTGGAGCGCTGGAGGACATTAACACACAGCTTCTTTTGG 2146
Qy      520 LysSerThrSerGlnGlnHiSleuThrLysSerAlaGlyLeuGlyLeuLysGlyTrpH 540
Db      2147 GCCTTTCACACAGCCAGACCACTCAGCAAGTCTGCTGGCCCTTAAGGCGCTGGC 2206
Qy      540 isSerAspIleLeuAlaProGlnThrSerThrProSerLeuThrSerSerTrpYrPhea 560
Db      2207 ACTGGATATCTTGGCCCCCGAACACTTAACCCCTTCCCTGACAGACAGCTGTATTGG 2266
Qy      560 laThrGluSerSerHiSpheYrSerAlaSerAlaIleYrGlyGlySerAlaSerYrS 580
Db      2267 CCACAGAGTCTCACACTTCTACTCTGCTCAGCCCATCTACGAGGCAATGCGCAAGTACT 2326
Qy      580 exrAlaYrSerCYsSerGlnLeuProThrCYsGlyAspGlnValYrSerValArgArg 600
Db      2327 CTGCCCTACAGCTGCACAGCTGCCCATCTTGGGAGAACCAAGCTATCTGTGCCAGGC 2386
Qy      600 rGlnLysProSerAspArgAlaAspSerArgArgSerTrpHiSgluGluSerProPheG 620
Db      2387 GCGAGAACCAAGTGCAGAGCTGACTCGCGCGGAGCTGGCATGAAGAGACCCCTTGG 2446
Qy      620 LuLysGlnPheLysArgArgSerCYsGlnMetGluPheGlyLysSerIleMetSerGlu 640
Db      2447 AAAAGCAGTTTAAACGACAGAGCTGCAGAAATGGAAATTTGGAGAGGCACTCATGTAGGA 2506
Qy      640 snArgSerArgGlnGluLeuGlyLysValGlySerGlnSerSerPheSerGlySerMetG 660
Db      2507 ACAGGTCACGGGAAAGCTGGGGAAAGTGGGCACTCAGCTTTCGGGCGAGCATGG 2566
Qy      660 LuIleIleGluValSer 665
Db      2567 AAATCATTCAGGCTCTCC 2583

```

RESULT 10

```

US-10-377-072-25
; Sequence 25, Application US/10377072
; Publication No. US20040009501A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals Inc.
; APPLICANT: Curtis, Kory A.J.
; APPLICANT: Logan, Thomas Joseph
; APPLICANT: Glucksmann, Maria A.
; APPLICANT: Meyers, Rachel E.

```

```

; APPLICANT: Williamson, Mark J.
; APPLICANT: Rudolph-Owen, Laura A.
; APPLICANT: Chun, Miyoung
; APPLICANT: Teal, Feng-Ying
; TITLE OF INVENTION: NOVEL 25869, 25934, 26335, 50365, 21117,
; TITLE OF INVENTION: 38697, 46508, 16816, 16839, 49937, 49931 AND 49933 MOLECULES
; TITLE OF INVENTION: AND USES THEREFOR
; FILE REFERENCE: MP103-0180NM1M
; CURRENT APPLICATION NUMBER: US/10/377,072
; PRIOR FILING DATE: 2003-02-27
; PRIOR APPLICATION NUMBER: US 09/895,860
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/215,370
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: US 09/723,806
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: US 60/187,455
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: US 09/843,297
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: US 60/199,801
; PRIOR FILING DATE: 2000-04-26
; PRIOR APPLICATION NUMBER: US 09/861,801
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: US 60/205,508
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: US 09/816,494
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/815,419
; PRIOR FILING DATE: 2001-03-22
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 3544
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (589)...(2586)
; US-10-377-072-25

Alignment Scores:
Pred. No.: 0
Score: 472.00
Percent Similarity: 99.40%
Best Local Similarity: 99.40%
Query Match: 70.98%
DB: 16
Length: 3544
Matches: 663
Conservative: 0
Mismatch: 2
Indels: 4
Gaps: 0

US-10-029-345A-109 (1-665) x US-10-377-072-25 (1-3544)
Qy      1 MetAlaHisGluMetIleGlyThrGlnIleValThrGluArgLeuValAlaLeuLeuGlu 20
Db      589 ATGGCCCATGAGATGATTTGGAACTCAAAATTTGTAAGAGAGTGTGGCTGCTGCGAAG 648
Qy      21 SerGlyThrGlnLysValLeuLeuLeuIleAspSerArgProPheValGluYrSntTrs 40
Db      649 AGTGAACGGAAAGAGTGTGCTTAATTTGATGAGCGGCAATTTGTGGAAATACATCATCC 708
Qy      41 HisIleLeuGluAlaIleAsnIleAsnCYsSerLysLeuMetLysArgArgLeuGln 60
Db      709 CACATTTTGGAGCCATTATATCACTCTCCAACTTATTAAGCGAAGGTTGCAGACAG 768
Qy      61 AspLysValLeuIleThrGluLeuIleGlnHisSerAlaLysHisLysValAspIleAsp 80
Db      769 GACAAAGTTTATTTACAGAGCTCATCCAGCATTCAGCGAAACATTAAGTTGACATTGAT 828
Qy      81 CYsSerGlnLysValValValYrAspGlnSerSerGlnAspValAlaSerLeuSerSer 100
Db      829 TGCAGTCAGAAAGGTGTATTTACATCAATCAAACTCCCAAGATGTTGCTTCTCTTCA 888
Qy      101 AspCYsPheLeuThrValLeuLeuGlyLysLeuGluLysSerPheAsnSerValHisLeu 120

```

```

Db      889  |||GACTGTTTCTGACTGCTACTTGGGTTAAACTGGAGAAGAGCTTCAACTCTGTTCACTCG 948
Qy      121  |||LeuAaGlyGlyPheAaGluPheSerArgCysPheProGlyLeuGlyGlyLeuSer 140
Db      949  |||CTTGCAAGGGGGGTTTCTGAGTTCCTCGTGTTCCTGGCTCTGTGAGAGGAAATTC 1008
Qy      141  |||ThrLeuValProThrCysIleSerGlnProCysLeuProValAlaAsnIleGlyProThr 160
Db      1009 |||ACTTATGCTCCCTACCTGCACTTCTCAGCTTGTCTTACCTGTTGCCAACATGGGCCAAC 1068
Qy      161  |||ArgIleLeuProAsnLeuTyrLeuGlyCysGlnArgAspValLeuAsnIleGlyLeu 180
Db      1069 |||CGAATTCCTCCCAATCTTTATCTTGTGCTGCCAGCGAGATGCTCTCAACAGAGCTGAT- 1127
Qy      181  |||-GlnGlnAsnGlyIleGlyTyrValLeuAsnAlaSerTyr-ThrCysProLysProAsp 200
Db      1128 |||GCAGCAGAAATGGGATGGTTATGTGTAAATGCCAGCAA-TACCTGTCCAAAGCCTGACT 1186
Qy      200  |||heIIeProGluSerHisPheLeuArgValProValAsnAspSerPheCysGluValIleL 220
Db      1187 |||TTATCCCGCAGTCTCAATTCCTGCTGCTGCTGTGAATACAGCTTTGTGAGAAATTT 1246
Qy      220  |||euProTrpLeuAspLysSerValAspPheIleGluLysAlaLysAsnGlyCysV 240
Db      1247 |||TCCCGTGGTTGGACAAATCAGTAGATTTCATTGAGAAAGCAAAGCCTCAATGATGTG 1306
Qy      240  |||alleuValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIle 260
Db      1307 |||TTCCTAGTGCATGTTTAACTGGGATCTCCGCTCCGCCACCATCGCTATGCTCATCA 1366
Qy      260  |||eLysArgMetAspMetSerLeuAspGluAlaTyrTrpPheValLysGluLysArgPro 280
Db      1367 |||TGAAGGAGATGGACATGCTTTAGATGAAGCTTACGATTGTGAAAGAAAAAGCCCTA 1426
Qy      280  |||hrIleSerProAsnPheAsnPheLeuGlyGlnLeuLeuAspTyrGluLysIleLys 300
Db      1427 |||CTATATCTCCAACTTCATTTTCTGGGCCAATCCTGGACATATGAGAAAGATTAA 1486
Qy      300  |||bngIlnThrGlyAlaSerGlyProLysSerLysLeuLysLeuLysIleGluLysPro 320
Db      1487 |||ACCAGACTGGACATCAGGGGCCAAAGAGCAAATCAAGCTGCTGACCTGGAGAGCAA 1546
Qy      320  |||bngIlnProValProAlaValSerGluGlyGlnLysSerGluThrProLeuSerPro 340
Db      1547 |||ATGAACCTTCTCCTGTCTGTCTAGAGGGTGAAGAGAAAGCCAGAGCCCTCAGTCC 1606
Qy      340  |||roCysAlaAspSerAlaThrSerGluAlaAlaGlyGlnArgProValHisProAlaSer 360
Db      1607 |||CCTGTGCCACTCTGCTACTCAGAGGCGAGAGAGCAAGGCCCGTGCATCCCGCAGCG 1666
Qy      360  |||aIProSerValProSerValGlnProSerLeuLeuGluAspSerProLeuValGlnAla 380
Db      1667 |||TCCCGCAGCTGCCAGCGTGAAGCGCTGCTGTTAAGACAGCCCGCTGTACAGGGCC 1726
Qy      380  |||euSerGlyLeuHisIleuSerAlaAspArgLeuGluAspSerAsnLysLeuLysArgSer 400
Db      1727 |||TCAGTGGGCTGACCTGTCCGAGACAGAGCTGGAACAGCAATAAAGCTCAAGCTTCT 1786
Qy      400  |||heSerLeuAspIleLysSerValSerTyrSerAlaSerMetAlaIleSerLeuHisGly 420
Db      1787 |||TCTCTCTGATATCAAAATCAGTTTCTATTCAGCCAGCATGGCAGCATCTTCATGGCT 1846
Qy      420  |||heSerSerSerGluAspAlaLeuGluTyrTyrLysProSerThrThrLeuAspGlyThr 440
Db      1847 |||TCTCCTCATCAGAAAGATGCTTTGGAATATCAAAACTTCCACTACTCTGATGGAGCA 1906
Qy      440  |||snLysLeuGlyGlnPheSerProValGlnGluLeuSerGluGlnThrProGluThrSer 460
Db      1907 |||ACAAGCTAATGACAGTCTCCCTGTTACAGAACTAATCGAGACAGATCTCCGAAACCA 1966
Qy      460  |||roAspLysGluGluAlaSerIleProLysLysLeuGlnThrAlaArgProSerAspSer 480

```

```

Db      1967 |||CTGATTAAGGAGGAGCCAGCATCCCAAGAGCTGACAGCCCGCAGGCTTGACAGACC 2026
Qy      480  |||InSerLysArgLeuHisSerValArgThrSerSerSerGlyThrAlaGlnArgSerLeu 500
Db      2027 |||AGAGCAAGGATTTGATTCGTGTCAGAACAGAGCAGTGGCAGCCCGCCAGAGGCTCC 2086
Qy      500  |||euSerProLeuHisIleArgSerGlySerValGluAspAsnTyrHisThrSerPheLeu 520
Db      2087 |||TATCTCCTGATTCGAGAGTGGAGAGCTGGAGAGCAATTAACACACACTTCTCTTTG 2146
Qy      520  |||LysSerThrSerGlnGlnHisIleuThrLysSerAlaGlyLeuGlyLeuLysGlyTyr 540
Db      2147 |||GCCTTCCACCGCCAGCGACACTCAGAGATCTGCTGCGGCTTAAAGGCTGGC 2206
Qy      540  |||IleSerAspIleLeuAlaProGlnThrSerThrProSerLeuThrSerSerTyrTrp 560
Db      2207 |||ACTCGATATCTTGGCCCCCAGACTCTTACCTTCCCTGACAGACCTGGTATTTTG 2266
Qy      560  |||aThrGluSerSerHisPheTyrSerAlaSerAlaIleTyrGlyGlySerAlaSerTyr 580
Db      2267 |||CCAGAGTCTCTCACTTCTACTGCTGCTCAGCCATCTTACGAGGAGGAGTCCAGTTACT 2326
Qy      580  |||eAlaTyrSerCysSerGlnLeuProThrCysGlyAspGlnValTyrSerValArgArg 600
Db      2327 |||CTGCTACAGCTGACAGCCAGCTGCCACTTGGGAGAGCAAGTCTATTTCTGTCCG 2386
Qy      600  |||xGlnLysProSerAspArgAlaAspSerArgSerTrpHisGluGluSerProPhe 620
Db      2387 |||GCCAAGACCAAGTACAGAGCTGACTCGCGCGGAGCTGGGATGAAGAGAGACCCCT 2446
Qy      620  |||LysGluPheLysArgArgSerCysGlnMetGluPheGlyGluSerIleMetSerGlu 640
Db      2447 |||AAAGCAGTTTAAAGCAGAACTGCAATGAAATTGGAAGACATCATGTCTGAGA 2506
Qy      640  |||snArgSerArgGluGluLeuGlyLysValGlySerGlnSerSerPheSerGlySerMet 660
Db      2507 |||ACAGGTCAAGGGAAGAGCTGGGAAAGTGGGAGTCAAGTCAAGTCTTTTGGGCA 2566
Qy      660  |||IleIleIleGluValSer 665
Db      2567 |||AAATCATTCAGCTCTCC 2583

RESULT 11
US-10-425-114-26234
; Sequence 26234, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 26234
; LENGTH: 3625
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB4119-028-H6_FLI
US-10-425-114-26234

Alignment Scores:
Pred. No.: 0
Score: 472.00
Percent Similarity: 99.40%
Best Local Similarity: 99.40%
Query Match: 70.98%
Length: 3625
Matches: 663
Conservative: 0
Mismatches: 2
Indels: 4

```


APPLICANT: ELLIOTT, Vicki S.; RAMKUNAR, Jayalaxmi
APPLICANT: YAO, Montique G.; BUREFORD, Nell
APPLICANT: WANG, Yumei E.; STEWART, Elizabeth A.
APPLICANT: GANDHI, Ameena R.; ARVIZU, Chandra S.
APPLICANT: LEE, Ernestine A.; HARALJA, April J.A.
APPLICANT: LU, Dying Aina M.; TRIBOULEY, Catherine M.
APPLICANT: GRIFFIN, Jennifer A.; BAUGHN, Mariah R.
APPLICANT: YUE, Henry; WARREN, Bridget A.
APPLICANT: NGUYEN, Daniel B.; CHAWLA, Narinder K.
APPLICANT: KEARNEY, Liam
TITLE OF INVENTION: PROTEIN PHOSPHATASES
FILE REFERENCE: PI-0173 PCT
CURRENT APPLICATION NUMBER: US/10/343,357
CURRENT FILING DATE: 2003-01-28
PRIOR APPLICATION NUMBER: PCT/US01/23716
PRIOR FILING DATE: 2001-07-26
PRIOR APPLICATION NUMBER: US 60/221,679
PRIOR FILING DATE: 2000-07-28
PRIOR APPLICATION NUMBER: US 60/223,272
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: US 60/224,309
PRIOR FILING DATE: 2000-08-10
PRIOR APPLICATION NUMBER: US 60/226,728
PRIOR FILING DATE: 2000-08-18
PRIOR APPLICATION NUMBER: US 60/229,254
PRIOR FILING DATE: 2000-08-30
PRIOR APPLICATION NUMBER: US 60/231,366
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 20
SOFTWARE: PERL Program
SEQ ID NO 17
LENGTH: 3766
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: m1sc feature
OTHER INFORMATION: Incyte ID No. US20040058341A1 7480570CBI
US-10-343-357-17

Alignment Scores:

Pred. No.: 0 Length: 3766
Score: 472.00 Matches: 663
Percent Similarity: 99.40% Conservative: 0
Best Local Similarity: 99.40% Mismatches: 2
Query Match: 70.98% Indels: 4
DB: 13 Gaps: 0

US-10-029-345a-109 (1-665) x US-10-343-357-17 (1-3766)

QY 1 MetAlAHlAGlUeUeTlleglYThrGlnlleValThGluArgleuValAlaleuleuGlu 20
DB 538 ATGGCCCATAGATGATGGAACTCAAAATTGTAACGAGAGGTGGCTGCTGCGTGA 597
QY 21 SerG1YThrG1uYrValleuLeu1leaPserArgProPheValG1uYrAanThrSer 40
DB 598 AGTGGAAACGGAAGAGTGCCTGATTAATGATGCGGCGCATTTGTGAATCAATACATCC 657
QY 41 HsIlleuGluAlaleuAnlleaenCySerlyleuWetlyArgrgleuGlnGln 60
DB 658 CACATTTTGGAGCCATTAATCACTGCTCCAGCTTATGAAGCGAAGTTGCAACG 717
QY 61 AspLyValleu1leThrguleu1leglnHlserAlalyshlalyValAaplleaP 80
DB 718 GACAAAGTGTAAATTCAGAGCTCATCCAGCATTCAGCGAACAATAGGTGAATTGAT 777
QY 81 CySerGlnlyrValValYrAspGlnSerSerGlnAspValAlaserleuSerSer 100
DB 778 TGCAGTCAGAGGTGTAGTTTACGATCAAGCTCCCAAGATTTGCTCTCTCTTCA 837
QY 101 AspCyPheleuthrValleuLeuGlylyleuGlnulysSerPheAnserValHlsleu 120
DB 838 GACGTGTTTCTCACTGTACTTCTGGGTAACTGAGAAAGACTTCAACTGTTCACTG 897

QY 121 LeuAlaGlylylPheAlaGluPheSerArgCyPheProGlyLeuCyGlnGlylylSer 140
DB 898 CTTCGACAGTGGGTTTGTGATGTTCTCTCTGTTTCCCTGCTGTGAAGAAATTC 957
QY 141 ThrleuValProThrCySllSerGlnProCySleuProValAlaAnlleGlyProThr 160
DB 958 ACTCTAGTCCCTACCTGACATTTCTGAGCTTCTGTTGCCAATGGGCCAACCC 1017
QY 161 ArgIleuProAnleuThrleuGlyCyGlnArgAspValleuAnlySgluleu1le 180
DB 1018 CCAATTCCTCCAAATCTTATATCTTGGCTGCCAGCGAAGTCTCAACAGAGCTGAT- 1076
QY 181 -GlnGlnaenglylleglYrValleuAnlaaserYr-ThrCySPolySProAsp 200
DB 1077 GAGCAGAAATGGGATTTGTTATGTGTAATGCCAGAA-TACCTGTCCAAAGCTGACT 1135
QY 200 HsIlleProGlnSerHlsPheleuArgValProValAsnAspSerPheCySglulyle 220
DB 1136 TTATCCCGAGTCTCATTTCTGCGTGCCTGTGAATGACAGCTTTGTGAGAAATTT 1195
QY 220 eurProThrleuAspLySerValAapPheIlleGlnulysAlalyValAlaserAnGlyCyAr 240
DB 1196 TCCCGGTGTGAGCAAAATGATGATTGATGAGAAACCAAGCCCAATGATGTG 1255
QY 240 allleuValHlsCySleuAlaGlylylSerArgSerAlaThrIleAlaIleATyrIlle 260
DB 1256 TTCTAGTCACGTGTTTACGTGGATCTCCCGCTCCGCAATCGCTTATCGCTTACATCA 1315
QY 260 eLlyeArgMetAspMetSerleuAspGluAlaTyArgPheVallySglulysArgProt 280
DB 1316 TGAAGAGATGAGCATGCTTTAGATGAGCTTACAGATTGTGAAAGAAAAAGACTTA 1375
QY 280 HsIlleSerProAnPheAnPheleuGlyGlnleuAnspYrGlyulysIllelyeAr 300
DB 1376 CTATATCTCCAACTTCAATTTTCTGGGCCCACTCTGACATGAGAAAGATTAAAG 1435
QY 300 sngGlnThrGlylaserGlyProLySerlyleuYleuHlsleuGlnlyePro 320
DB 1436 ACCAGCTGAGATCAGGAGCCCAAGAGCAAACTCAAGCTGCTGCACCTGAGAACCA 1495
QY 320 sngGlnProValProAlaValaserGlyGlyGlnlyeSerGlnThrProleuSerPro 340
DB 1496 ATGAACTCTCCCTGCTGCTCAGAGGAGGAGCAAGAAAGCGAGAGCCCTCAGTCCAC 1555
QY 340 rOcYeaAlaPserAlaThrSerGluAlaAlaGlyGlnArgProValHlsProAlaSerY 360
DB 1556 CCTGTGCCACTGCTCTACCTCAGAGGACAGCAAGAGCCCGCTGCAATCCGCCAGCC 1615
QY 360 a1ProSerValProSerValGlnProSerleuGlnuAspSerProleuValGlnAla 380
DB 1616 TGCACAGCTGCCAGCGTGCAGCGCTCGCTGTAAAGACAGCCGCTGGTACAGGCGC 1675
QY 380 eusSerGlyleuHlsleuSerAlaAspArgleuGlnuAspSerAnlySleuLyArgSerP 400
DB 1676 TCAGTGGCTGCACCTGTCGCGAGACAGGCTGGAAGACGAATAGCTCAAGCGTTCC 1735
QY 400 HsSerleuAspIllelyeSerValSerTySerAlaserMetAlaAlaserleuHlsGlyP 420
DB 1736 TCTCTCTGATATCAAAATGATTAATTCAGCGCATGCGAGATCTTACATGAGCT 1795
QY 420 HsSerSerGlnuAspAlaleuGluTyTylyAspSerThrThrleuAspGlyThra 440
DB 1796 TCTCTCATCAGAAAGTCTTGGAAATCTACCAACCTTCACTACTGATGGAGCA 1855
QY 440 smllyeLeuCySglnPheSerProValGlnGlyleuSerGlnGlnThrProGlnThrSerP 460
DB 1856 ACAAGCTATGCCAGTCTCCCGTTCAGAAATATCGAAGAGACTCCCGAAACCAAGTC 1915
QY 460 rOAspLyGlnGluAlaSerIlleProLylyleuGlnThrAlaArgProSerAspSerG 480
DB 1916 CTGATTAAGAGAGAGCCACATCCCAAGAGCTGCAGAGCCGCTTCAAGACAGCC 1975
QY 480 lnseryArgleuHlsSerValArgThrSerSerSerGlyThrAlaGlnArgSerleu 500


```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2429
; LENGTH: 1916
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-108-260A-2429

Alignment Scores:
Pred. No.: 1,6e-169 Length: 1916
Score: 178.00 Matches: 278
Percent Similarity: 99.29% Conservative: 0
Best Local Similarity: 99.29% Mismatches: 1
Query Match: 26.77% Indels: 2
DB: 16 Gaps: 0

US-10-029-345A-109 (1-665) x US-10-108-260A-2429 (1-1916)

QY 387 AAlaSPaRgLeuGluuAPSerAaNLyLeuLYaRgSerPheSerLeuApIleLYsSer 406
Db 142 GCAGACAGGCTGGAGAGACAGCAATAGCTCAGCTCTCTCTGATATCAATCA 201

QY ValSerTYrSerAlaSerMetAlaLaserLeuHIGLYPheSerSerSerGluuAPAla 426
Db 202 GTTTCATATTCAAGCCAGCATGGAGCATCTTCTACATAGCTTCTCTCATCAGAAATGCT 261

QY 427 LeuGluTYrTYrLYsPProSerThrThrLeuAPGLYThraNLyLeuCYsGlnPheSer 446
Db 262 TTGGAAATACACAAACCTTCCACTACTGTGATGGAGCAAGCATATGCCAGTTCTCC 321

QY 447 ProValGlnGluLeuSerGluGlnThrProGlnThrSerProApLYsGluGluAlaSer 466
Db 322 CCGTTTCAGGAACATATCGAGAGCACTCCGAAACAGTCGATATAGAGGAGGACACG 381

QY 467 lLeProLYsLYsLeuGlnThraLaArgProSerAPSerGlnSerLYsAArgLeuHISer 486
Db 382 ATCCCAAGAGAGCTGAGAGCCGCCAGCCCTTCAGAGACCAAGCAAGCATTCGATTCG 441

QY 487 ValaRThSerSerSerGlyThraLaGlnaRgSerLeuLeuSerProLeuHISaRgSer 506
Db 442 GTCAGAACCAAGCGAGCTGAGCAAGCCAGAGGTCCTTTATCTTCATCGCATGGAAGT 501

QY 507 GLYSerValGluAPaNLyRThSerThrPheLeuPheGlyLeuSerThrSerGlnG 526
Db 502 GGGAGGCTGAGAGCAATTAACAACACACTTCTTTTGCGCTTCCAT-CAGCGAGCA 560

QY 526 nHISLeuThrLYsSerAlaGlyLeuGlyLeuLYsGlyTPHISerAPAlaLeuAlaPr 546
Db 561 GCACCTCAAGAGTCTGCTGGCTGGCTTAAGGCTTGCACTCGGATATCTTGAGCCCC 620

QY 546 oGlnThrSerThrProSerLeuThSerSerTrpTYrPheAlaThRgLeuSerSerHISpH 566
Db 621 CCAGACCTTACCCCTTCCCTGACCAAGCGAGCTGATTTTGCACAGAGCTCTCACACT 680

QY 566 eTYrSerAlaSerAlaIleTYrGlyGlySerAlaSerTYrSerAlaTYrSerCYsSerG 586
Db 681 CTRACTGCTTACGCATTTACGAGAGCGAGTCCAGTACTGCTGCTACAGTGCAGCA 740

QY 586 nLeuProThrCYsGlyAPSPAlnValTYrSerValaRgaRgaGlnLYsProSerAPaR 606
Db 741 GCTGCCCACTTGCAGAGACCAAGTCTATCTGCGCAGGCGCGCAAGGCCAAGAGCAAG 800

QY 606 gAlaAPSerArGaRgSerTrpHIGlyGluSerProPheGluLYsGlnPheLYsARAR 626
Db 801 AGCTACTGCGCGCGAGCTGAGCAAGAGAGAGCCCTTTGAAAGCACTTAAACGAG 860

QY 626 gSerCYsGlnMetGluPheGlyGluSerlLeuSerGluuAPaNLyAPaRgGluGluLe 646
Db 861 AAGCTGCCAAATGAAATTGGAGAGAGCATCTGCTCAAGAGAACAGGTACCGGAAAGACT 920

QY 646 uGlyLYsValGlySerGlnSerSerPheSerGlySerMetGluIlelGluValaSer 665
Db 921 GGGGAAAGTGGGAGTCAAGTCTAGCTTTTCGGGCGAGCATGAAATCATTAAGGTCTCC 978

```

```

RESULT 17
US-10-005-858-1
; Sequence 1, Application US/10005858
; Publication No. US20020116729A1
; GENERAL INFORMATION:
; APPLICANT: Allen, Keith D.
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING NTPPI
; FILE REFERENCE: R-690
; CURRENT APPLICATION NUMBER: US/10/005,858
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: US 60/251,802
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2453
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-005-858-1

Alignment Scores:
Pred. No.: 1.1e-09 Length: 2453
Score: 20.00 Matches: 20
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.01% Indels: 0
DB: 14 Gaps: 0

US-10-029-345A-109 (1-665) x US-10-005-858-1 (1-2453)

QY 242 ValHISCYsLeuAlaGlyLYsEArSerAPSerAlaThrlLeAlaIleAlaTYrLeMetLYs 261
Db 827 GTTCACTGCTGCTGGCATCTGCTCTGCAACCAATGCGGTACATCATGATAA 886

RESULT 18
US-10-220-120-28
; Sequence 28, Application US/10220120
; Publication No. US20040408253A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: PANZER, Scott R.
; APPLICANT: SPIRO, Peter A.
; APPLICANT: BANVILLE, Steven C.
; APPLICANT: SHAH, Purvi
; APPLICANT: CHALUP, Michael S.
; APPLICANT: CHANG, Simon C.
; APPLICANT: CHEN, Alice
; APPLICANT: D'SA, Steven A.
; APPLICANT: AMSHEY, Stefan
; APPLICANT: DAHL, Christopher R.
; APPLICANT: DAM, Tam C.
; APPLICANT: DANIELS, Susan E.
; APPLICANT: DUPOUR, Gerard E.
; APPLICANT: FLORES, Vincent
; APPLICANT: FONG, Willy T.
; APPLICANT: GREENMALT, Lila B.
; APPLICANT: HILLMAN, Jennifer L.
; APPLICANT: JONES, Antissa L.
; APPLICANT: LIU, Tommy F.
; APPLICANT: ROSEBERRY, Ann M.
; APPLICANT: ROSEN, Bruce H.
; APPLICANT: RUSSO, Frank D.
; APPLICANT: STOCKREHER, Theresa K.
; APPLICANT: DAFPO, Abel
; APPLICANT: WRIGHT, Rachel J.
; APPLICANT: YAP, Pierre E.
; APPLICANT: YU, Jimmy Y.
; APPLICANT: BRADLEY, Diana L.
; APPLICANT: BRATCHER, Shawn R.
; APPLICANT: CHEN, Wensheng
; APPLICANT: COHEN, Howard J.
; APPLICANT: HODGSON, David M.

```



```
; APPLICANT: LINCOLN, Stephen E.
; APPLICANT: JACKSON, Stuart
; TITLE OF INVENTION: MOLECULES FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: FT-1113 PCT
; CURRENT APPLICATION NUMBER: US/10/220,120
; PRIOR FILING DATE: 2002-08-26
; PRIOR APPLICATION NUMBER: 60/184,777; 60/184,797; 60/184,698; 60/184,770; 60/184,774;
; 60/184,693; 60/184,771; 60/184,813; 60/184,773; 60/184,776;
; 60/184,769; 60/184,766; 60/184,837; 60/184,697; 60/184,841;
; 60/184,772; 60/185,213; 60/185,216; 60/204,863; 60/205,221;
; 60/204,815; 60/203,785; 60/204,821; 60/204,908; 60/204,226;
; 60/204,525; 60/205,285; 60/205,232; 60/205,323; 60/205,287;
; 60/205,324; 60/205,286
; PRIOR FILING DATE: 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24;
; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24;
; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24; 2000-02-24;
; 2000-05-17; 2000-05-12; 2000-02-24; 2000-02-24; 2000-05-17;
; 2000-05-16; 2000-05-12; 2000-05-16; 2000-05-16; 2000-05-15;
; 2000-05-17; 2000-05-17; 2000-05-16; 2000-05-17; 2000-05-17;
; NUMBER OF SEQ ID NOS: 422
; SOFTWARE: PERL Program
; SEQ ID NO 28
; LENGTH: 2476
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20040048253A1 LI:1085273.2:2000MAY01
US-10-029-345A-109 (1-665) x US-10-220-120-28 (1-2476)

Alignment Scores:
Pred. No.: 1,11e-09 Length: 2476
Score: 20.00 Matches: 20
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.01% Indels: 0
DB: 13 Gaps: 0

Qy 242 ValHhCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIleMetLys 261
Db 933 GTCCATCTGCTGGCTGGCATCTCCGCTCTGCCACCATGCCATGCCCTTACATCATGAGG 992

RESULT 19
US-09-925-299-69
; Sequence 69, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 69
; LENGTH: 1848
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: n equals a,t,g, or c
; APPLICANT: Bret P. Monia
; APPLICANT: Lex M. Cowseart
; APPLICANT: Kenneth W. Doble
; TITLE OF INVENTION: ANTISENSE MODULATION OF DUAL SPECIFIC PHOSPHATASE 6 EXPRESSION
```

```
; LOCATION: (1847)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-69

Alignment Scores:
Pred. No.: 0.104 Length: 1848
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: 9 Gaps: 0

US-10-029-345A-109 (1-665) x US-09-925-299-69 (1-1848)

Qy 240 ValLeuValHhCysLeuAlaGlyIleSerArgSer 251
Db 290 GTCTTGATCAATGCTGGCTGGCATTCACCCGCTCA 325

RESULT 20
US-09-925-299-69
; Sequence 69, Application US/09925299
; Publication No. US20030040617A9
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 69
; LENGTH: 1848
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: (1761)
; LOCATION: (1761)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc.feature
; LOCATION: (1844)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc.feature
; LOCATION: (1847)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-69

Alignment Scores:
Pred. No.: 0.104 Length: 1848
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: 10 Gaps: 0

US-10-029-345A-109 (1-665) x US-09-925-299-69 (1-1848)

Qy 240 ValLeuValHhCysLeuAlaGlyIleSerArgSer 251
Db 290 GTCTTGATCAATGCTGGCTGGCATTCACCCGCTCA 325

RESULT 21
US-10-199-221-11
; Sequence 11, Application US/10199221
; Publication No. US20040014048A1
; GENERAL INFORMATION:
; APPLICANT: Bret P. Monia
; APPLICANT: Lex M. Cowseart
; APPLICANT: Kenneth W. Doble
; TITLE OF INVENTION: ANTISENSE MODULATION OF DUAL SPECIFIC PHOSPHATASE 6 EXPRESSION
```



```

; FILE REFERENCE: PTS-0009
; CURRENT APPLICATION NUMBER: US/10/199,221
; CURRENT FILING DATE: 2002-07-18
; NUMBER OF SEQ ID NOS: 101
; SEQ ID NO 11
; LENGTH: 1952
; TYPE: DNA
; ORGANISM: H. sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (352)...(1059)
US-10-199-221-11

Alignment Scores:
Pred. No.: 0.109      Length: 1952
Score: 12.00      Matches: 12
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 1.80%      Indels: 0
DB: 16      Gaps: 0

US-10-029-345A-109 (1-665) x US-10-199-221-11 (1-1952)

QY 240 ValLeuValHisCysLeuAlaGlyIleSerArgSer 251
Db 778 GCTTGGTACATGCTTGCTGGCATTAAGCCGCTCA 813

RESULT 22
US-10-072-012-265
; Sequence 265, Application US/10072012
; Publication No. US20040033493A1
; GENERAL INFORMATION:
; APPLICANT: Tcherenev, Velizar
; APPLICANT: Spytek, Kimberly
; APPLICANT: Zehusen, Bryan
; APPLICANT: Patturajan, Meena
; APPLICANT: Shimkets, Richard
; APPLICANT: Li, Li
; APPLICANT: Gangolli, Esha
; APPLICANT: Padigaru, Maralidhara
; APPLICANT: Anderson, David W.
; APPLICANT: Rastelli, Luca
; APPLICANT: Miller, Charles E.
; APPLICANT: Gerlach, Valerie
; APPLICANT: Taupier Jr, Raymond J.
; APPLICANT: Gusev, Vladimир Y.
; APPLICANT: Colman, Steven D.
; APPLICANT: Wolenc, Adam R.
; APPLICANT: Pena, Carol E. A
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Grose, William M.
; APPLICANT: Alsobrook II, John P.
; APPLICANT: Lepley, Denise M.
; APPLICANT: Rieger, Daniel X.
; APPLICANT: Burgess, Catherine E.
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-258
; CURRENT APPLICATION NUMBER: US/10/072,012
; CURRENT FILING DATE: 2002-01-31
; PRIOR APPLICATION NUMBER: 60/265,102
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/265,514
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,517
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,412
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/265,395
; PRIOR FILING DATE: 2001-01-31
; PRIOR APPLICATION NUMBER: 60/266,406
; PRIOR FILING DATE: 2001-02-02
; PRIOR APPLICATION NUMBER: 60/266,767
```

```

; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/267,057
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/266,975
; PRIOR FILING DATE: 2001-02-07
; PRIOR APPLICATION NUMBER: 60/267,459
; PRIOR FILING DATE: 2001-02-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1391
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 265
; LENGTH: 2039
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-072-012-265

Alignment Scores:
Pred. No.: 0.114      Length: 2039
Score: 12.00      Matches: 12
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 1.80%      Indels: 0
DB: 13      Gaps: 0

US-10-029-345A-109 (1-665) x US-10-072-012-265 (1-2039)

QY 281 lIeSerProAsnPhaeAsnPhaeLeuGlyGlnLeuLeu 292
Db 952 ATCTGCCCACTTCAACTTCTGCGCAGCTGCTG 987

RESULT 23
US-10-152-319A-2041
; Sequence 2041, Application US/10152319A
; Publication No. US20040072160A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Higgs, Brandon
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5089-US
; CURRENT APPLICATION NUMBER: US/10/152,319A
; CURRENT FILING DATE: 2002-05-22
; PRIOR APPLICATION NUMBER: US 60/292,335
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/297,523
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,925
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,810
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/303,807
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/303,808
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/315,047
; PRIOR FILING DATE: 2001-08-28
; PRIOR APPLICATION NUMBER: US 60/324,928
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/330,867
; PRIOR FILING DATE: 2001-11-01
; PRIOR APPLICATION NUMBER: US 60/330,462
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 2221
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2041
; LENGTH: 2104
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
```

```
; OTHER INFORMATION: Genbank Accession No. NM_053883
US-10-152-319A-2041

Alignment Scores:
Pred. No.: 0.117      Length: 2104
Score: 12.00          Matches: 12
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 1.80%      Indels: 0
DB: 12                Gaps: 0

US-10-029-345A-109 (1-665) x US-10-152-319A-2041 (1-2104)
QY 240 ValLeuValHicCySeuAaIagIyIleSeArGser 251
DB 1225 GTCCTGTGATTCCTTGCGGCGCATCAGCCGCTCC 1260

RESULT 24
US-10-388-934-654
; Sequence 654, Application US/10388934
; Publication No. US20040005547A1
; GENERAL INFORMATION:
; APPLICANT: Boeser, Franziska
; APPLICANT: Suter-Dick, Laura
; APPLICANT: Wolf, Detlef
; TITLE OF INVENTION: BIOMARKERS AND EXPRESSION PROFILES FOR TOXICOLOGY
; FILE REFERENCE: 21199
; CURRENT APPLICATION NUMBER: US/10/388,934
; CURRENT FILING DATE: 2003-03-14
; PRIOR APPLICATION NUMBER: 02005336.9
; PRIOR FILING DATE: 2002-03-14
; PRIOR APPLICATION NUMBER: 02015657.6
; PRIOR FILING DATE: 2002-07-17
; NUMBER OF SEQ ID NOS: 862
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 654
; LENGTH: 2104
; TYPE: DNA
; ORGANISM: Rattus norvegicus (No. US20040005547A1way rat)
US-10-388-934-654

Alignment Scores:
Pred. No.: 0.117      Length: 2104
Score: 12.00          Matches: 12
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 1.80%      Indels: 0
DB: 16                Gaps: 0

US-10-029-345A-109 (1-665) x US-10-388-934-654 (1-2104)
QY 240 ValLeuValHicCySeuAaIagIyIleSeArGser 251
DB 1225 GTCCTGTGATTCCTTGCGGCGCATCAGCCGCTCC 1260

RESULT 25
US-10-191-803-270
; Sequence 270, Application US/10191803
; Publication No. US20040014040A1
; GENERAL INFORMATION:
; APPLICANT: MENDRICK, Donna
; APPLICANT: PORTER, Mark
; APPLICANT: JOHNSON, Kory
; APPLICANT: HIGGS, Brandon
; APPLICANT: CASTLE, Arthur
; APPLICANT: BLASHOF, Michael
; TITLE OF INVENTION: Cardiotocxin Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5090US
; CURRENT APPLICATION NUMBER: US/10/191,803
; CURRENT FILING DATE: 2002-07-10
; PRIOR APPLICATION NUMBER: US 60/303,819
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/305,623
```

```
; PRIOR FILING DATE: 2001-07-17
; PRIOR APPLICATION NUMBER: US 60/369,351
; PRIOR FILING DATE: 2002-04-03
; PRIOR APPLICATION NUMBER: US 60/377,611
; PRIOR FILING DATE: 2002-05-06
; NUMBER OF SEQ ID NOS: 1140
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 270
; LENGTH: 2104
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20040014040A1 NM_053883
US-10-191-803-270

Alignment Scores:
Pred. No.: 0.117      Length: 2104
Score: 12.00          Matches: 12
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 1.80%      Indels: 0
DB: 16                Gaps: 0

US-10-029-345A-109 (1-665) x US-10-191-803-270 (1-2104)
QY 240 ValLeuValHicCySeuAaIagIyIleSeArGser 251
DB 1225 GTCCTGTGATTCCTTGCGGCGCATCAGCCGCTCC 1260

RESULT 26
US-10-305-720-1135
; Sequence 1135, Application US/10305720
; Publication No. US20040010136A1
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice K.; Seilhamer, Jeffrey J.
; TITLE OF INVENTION: Composition for the Detection of Signaling Pathway Gene Expressi
; FILE REFERENCE: PA-0002-1 CON
; CURRENT APPLICATION NUMBER: US/10/305,720
; CURRENT FILING DATE: 2002-11-26
; PRIOR APPLICATION NUMBER: 09/016,434
; PRIOR FILING DATE: 1998-01-30
; NUMBER OF SEQ ID NOS: 1490
; SOFTWARE: PERL Program
; SEQ ID NO 1135
; LENGTH: 2109
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Genbank ID No. US20040010136A1 g1418933
US-10-305-720-1135

Alignment Scores:
Pred. No.: 0.117      Length: 2109
Score: 12.00          Matches: 12
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 1.80%      Indels: 0
DB: 16                Gaps: 0

US-10-029-345A-109 (1-665) x US-10-305-720-1135 (1-2109)
QY 240 ValLeuValHicCySeuAaIagIyIleSeArGser 251
DB 1216 GTCCTGTGATTCCTTGCGGCGCATCAGCCGCTCA 1251

RESULT 27
US-10-641-643-946
; Sequence 946, Application US/10641643
; Publication No. US20040077003A1
; GENERAL INFORMATION:
; APPLICANT: Cocks, Benjamin G.
; APPLICANT: Susan G. Stuart
```

Jeffrey J. Seilhamer
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL
GENE EXPRESSION
NUMBER OF SEQUENCES: 1508
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/641,643
FILING DATE: 14-Aug-2003
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: <Unknown>
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Zeller, Karen J.
REGISTRATION NUMBER: 37,071
REFERENCE/DOCKET NUMBER: PA-0001 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 946:
SEQUENCE CHARACTERISTICS:
LENGTH: 2109 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GENBANK
CLONE: g1418933
SEQUENCE DESCRIPTION: SEQ ID NO: 946 :
US-10-641-643-946
Alignment Scores:
Pred. No.: 0.117 Length: 2109
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: 17 Gaps: 0
US-10-029-345A-109 (1-665) x US-10-641-643-946 (1-2109)
QY 240 ValLeuValHISCySeLeuAlaGlyIleSeRArgSer 251
Db 1216 GTCCTGTGACATTGCTTGCTGGCATTTAGCCGCTCA 1251
RESULT 28
US-10-342-887-698
Sequence 698, Application US/10342887
GENERAL INFORMATION:
APPLICANT: Dai, Hongyue
APPLICANT: He, Yudong
APPLICANT: Linsley, Peter S.
APPLICANT: Mao, Mao
APPLICANT: Roberts, Christopher J.
APPLICANT: Van 't Veer, Laura Johanna
APPLICANT: Bernards, Rene
TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
FILE REFERENCE: 9301-188-999
CURRENT APPLICATION NUMBER: US/10/342,887
CURRENT FILING DATE: 2003-01-15

PRIOR APPLICATION NUMBER: 60/298,918
PRIOR FILING DATE: 2001-06-18
PRIOR APPLICATION NUMBER: 60/380,710
PRIOR FILING DATE: 2002-05-14
PRIOR APPLICATION NUMBER: 10/172,118
PRIOR FILING DATE: 2002-06-14
NUMBER OF SEQ ID NOS: 2699
SEQ ID NO 698
LENGTH: 2390
TYPE: DNA
ORGANISM: Homo sapiens
US-10-342-887-698
Alignment Scores:
Pred. No.: 0.131 Length: 2390
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: 13 Gaps: 0
US-10-029-345A-109 (1-665) x US-10-172-118-698 (1-2390)
QY 240 ValLeuValHISCySeLeuAlaGlyIleSeRArgSer 251
Db 1216 GTCCTGTGACATTGCTTGCTGGCATTTAGCCGCTCA 1251
RESULT 30
US-10-175-523-183
Sequence 183, Application US/10175523
Alignment Scores:
Pred. No.: 0.131 Length: 2390
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: 13 Gaps: 0
US-10-029-345A-109 (1-665) x US-10-172-118-698 (1-2390)
QY 240 ValLeuValHISCySeLeuAlaGlyIleSeRArgSer 251
Db 1216 GTCCTGTGACATTGCTTGCTGGCATTTAGCCGCTCA 1251
RESULT 30
US-10-175-523-183
Sequence 183, Application US/10175523

Publication No. US20030096264A1
GENERAL INFORMATION:
APPLICANT: Brockman, Jeffrey
APPLICANT: Evans, David
APPLICANT: Hook, Derek
APPLICANT: Klimczak, Leszek
APPLICANT: Laeng, Pascal
APPLICANT: Paleyeyman, Michael
APPLICANT: Rajan, Prithi
TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPTS)
FILE REFERENCE: 3235/13795-US3
CURRENT APPLICATION NUMBER: US/10/175,523
CURRENT FILING DATE: 2002-06-18
PRIOR APPLICATION NUMBER: US 60/299,151
PRIOR FILING DATE: 2001-06-18
PRIOR APPLICATION NUMBER: US 60/317,828
PRIOR FILING DATE: 2001-09-07
PRIOR APPLICATION NUMBER: US 60/325,150
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US 60/333,047
PRIOR FILING DATE: 2001-11-14
PRIOR APPLICATION NUMBER: US 60/349,936
PRIOR FILING DATE: 2002-01-18
PRIOR APPLICATION NUMBER: US 60/361,834
PRIOR FILING DATE: 2002-03-04
NUMBER OF SEQ ID NOS: 197
SOFTWARE: PatentIn version 3.1
SEQ ID NO 183
LENGTH: 2390
TYPE: DNA
ORGANISM: Homo sapiens
US-10-175-523-183

Alignment Scores:
Pred. No.: 0.131 Length: 2390
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: 15 Gaps: -0

US-10-029-345A-109 (1-665) x US-10-175-523-183 (1-2390)
QY 240 ValLeuValHicCyLeuAlaGlyIleSerArgSer 251
DB 1216 GTCCTGTACATTCCTGCTGCGCATTAAGCCGCTCA 1251

RESULT 31
US-10-199-221-4
Sequence 4, Application US/10199221
Publication No. US20040014048A1
GENERAL INFORMATION:
APPLICANT: Brett P. Monia
APPLICANT: Lex M. Coweart
APPLICANT: Kenneth W. Dobie
TITLE OF INVENTION: ANTISENSE MODULATION OF DUAL SPECIFIC PHOSPHATASE 6 EXPRESSION
FILE REFERENCE: FTS-0009
CURRENT APPLICATION NUMBER: US/10/199,221
CURRENT FILING DATE: 2002-07-18
NUMBER OF SEQ ID NOS: 101
SEQ ID NO 4
LENGTH: 2390
TYPE: DNA
ORGANISM: H. sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (352)...(1497)
US-10-199-221-4
Alignment Scores:
Pred. No.: 0.131 Length: 2390
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: 15 Gaps: -0

Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: 16 Gaps: 0

US-10-029-345A-109 (1-665) x US-10-199-221-4 (1-2390)

QY 240 ValLeuValHicCyLeuAlaGlyIleSerArgSer 251
DB 1216 GTCCTGTACATTCCTGCTGCGCATTAAGCCGCTCA 1251

RESULT 32
US-10-717-597-4
Sequence 4, Application US/10717597
Publication No. US20040110221A1
GENERAL INFORMATION:
APPLICANT: Wyeth
APPLICANT: Burczynski, Michael E.
APPLICANT: Twine, Natalie C.
APPLICANT: Dornier, Andrew J.
APPLICANT: Trepichio, William L.
APPLICANT: Slonim, Donna K.
APPLICANT: Stover, Jennifer A.
TITLE OF INVENTION: METHODS FOR DIAGNOSING RCC AND OTHER SOLID TUMORS
FILE REFERENCE: AM101080L
CURRENT APPLICATION NUMBER: US/10/717,597
CURRENT FILING DATE: 2003-11-21
PRIOR APPLICATION NUMBER: US 60/459,782
PRIOR FILING DATE: 2003-04-03
PRIOR APPLICATION NUMBER: US 60/427,982
PRIOR FILING DATE: 2002-11-21
NUMBER OF SEQ ID NOS: 4904
SOFTWARE: PatentIn version 3.2
SEQ ID NO 4
LENGTH: 2390
TYPE: DNA
ORGANISM: Homo sapiens
US-10-717-597-4

Alignment Scores:
Pred. No.: 0.131 Length: 2390
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: 17 Gaps: 0

US-10-029-345A-109 (1-665) x US-10-717-597-4 (1-2390)

QY 240 ValLeuValHicCyLeuAlaGlyIleSerArgSer 251
DB 1216 GTCCTGTACATTCCTGCTGCGCATTAAGCCGCTCA 1251

RESULT 33
US-10-133-937-48
Sequence 48, Application US/10133937
Publication No. US20030207278A1
GENERAL INFORMATION:
APPLICANT: Khan, Javed
APPLICANT: Ringner, Markus
APPLICANT: Peterson, Carsten
APPLICANT: Melzer, Paul
TITLE OF INVENTION: METHODS FOR ANALYZING HIGH DIMENSIONAL DATA FOR CLASSIFYING, DIAGNOSING, PROGNOSTICATING, AND/OR PREDICTING DISEASES AND
FILE REFERENCE: 11613.56US01
CURRENT APPLICATION NUMBER: US/10/133,937
CURRENT FILING DATE: 2002-11-04
NUMBER OF SEQ ID NOS: 99
SOFTWARE: PatentIn version 3.1
SEQ ID NO 48
LENGTH: 2649
TYPE: DNA

```
/ ORGANISM: Homo sapiens
US-10-133-937-48

Alignment Scores:
Pred. No.: 0.144 Length: 2649
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: 16 Gaps: 0

US-10-029-345A-109 (1-665) x US-10-133-937-48 (1-2649)
QY 240 ValieuValHiscysleuA1aglyI1eserA9ser 251
Db 1153 GTCCTGGTACATTCCTTGGCTGGCATTTAGCCGCTCA 1188

RESULT 34
US-10-159-563-48
/ Sequence 48, Application US/10159563
/ Publication No. US20040009154A1
/ GENERAL INFORMATION:
/ APPLICANT: Khan, Javed
/ APPLICANT: Ringner, Markus
/ APPLICANT: Peterson, Carsten
/ APPLICANT: Meltzer, Paul
/ TITLE OF INVENTION: SELECTIONS OF GENES AND METHODS OF USING THE SAME FOR
/ TITLE OF INVENTION: DIAGNOSIS AND FOR TARGETING THE THERAPY OF SELECT CANCERS
/ FILE REFERENCE: 11613.56US11
/ CURRENT APPLICATION NUMBER: US/10/159,563
/ CURRENT FILING DATE: 2002-12-09
/ PRIOR APPLICATION NUMBER: US 10/133,937
/ PRIOR FILING DATE: 2002-04-25
/ NUMBER OF SEQ ID NOS: 444
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 48
/ LENGTH: 2649
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-159-563-48

Alignment Scores:
Pred. No.: 0.144 Length: 2649
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: 16 Gaps: 0

US-10-029-345A-109 (1-665) x US-10-159-563-48 (1-2649)
QY 240 ValieuValHiscysleuA1aglyI1eserA9ser 251
Db 1153 GTCCTGGTACATTCCTTGGCTGGCATTTAGCCGCTCA 1188

RESULT 35
US-10-184-832-1
/ Sequence 1, Application US/10184832
/ Publication No. US20030022857A1
/ GENERAL INFORMATION:
/ APPLICANT: Xu et al.
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
/ TITLE OF INVENTION: TREATMENT OF BODY WEIGHT DISORDERS, INCLUDING OBESITY
/ FILE REFERENCE: MP12001-056P1NM
/ CURRENT APPLICATION NUMBER: US/10/184,832
/ CURRENT FILING DATE: 2002-06-28
/ PRIOR APPLICATION NUMBER: 60/303,250
/ PRIOR FILING DATE: 2001-07-05
/ NUMBER OF SEQ ID NOS: 6
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 1
/ LENGTH: 2650
/ TYPE: DNA
```

```
/ ORGANISM: Mus musculus
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (280) ... (1422)
US-10-184-832-1

Alignment Scores:
Pred. No.: 0.144 Length: 2650
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: 15 Gaps: 0

US-10-029-345A-109 (1-665) x US-10-184-832-1 (1-2650)
QY 240 ValieuValHiscysleuA1aglyI1eserA9ser 251
Db 1144 GTCCTGGTACATTCCTTGGCAGTATTCAGCCGCTCT 1179

RESULT 36
US-09-971-392-69
/ Sequence 69, Application US/09971392
/ Publication No. US20030134283A1
/ GENERAL INFORMATION:
/ APPLICANT: Peterson, David P.
/ APPLICANT: Pearson, Cecelia I.
/ APPLICANT: Cocks, Benjamin G.
/ TITLE OF INVENTION: GENES REGULATED IN DENDRITIC CELL DIFFERENTIATION
/ FILE REFERENCE: PA-0029 US
/ CURRENT APPLICATION NUMBER: US/09/971,392
/ CURRENT FILING DATE: 2001-10-03
/ PRIOR APPLICATION NUMBER: 60/237,652
/ PRIOR FILING DATE: 2000-10-03
/ NUMBER OF SEQ ID NOS: 260
/ SOFTWARE: PERL Program
/ SEQ ID NO 69
/ LENGTH: 2792
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc feature
/ OTHER INFORMATION: Template ID: 149914.21
/ NAME/KEY: unsure
/ LOCATION: 2789
/ OTHER INFORMATION: a, t, c, g, or other
US-09-971-392-69

Alignment Scores:
Pred. No.: 0.151 Length: 2792
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: 10 Gaps: 0

US-10-029-345A-109 (1-665) x US-09-971-392-69 (1-2792)
QY 240 ValieuValHiscysleuA1aglyI1eserA9ser 251
Db 1314 GTCCTGGTACATTCCTTGGCTGGCATTTAGCCGCTCA 1349

RESULT 37
US-10-044-090-330
/ Sequence 330, Application US/10044090
/ Publication No. US20020137081A1
/ GENERAL INFORMATION:
/ APPLICANT: Olga Bandman
/ TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
/ FILE REFERENCE: PA-0028 US
/ CURRENT APPLICATION NUMBER: US/10/044,090
/ CURRENT FILING DATE: 2002-01-09
/ NUMBER OF SEQ ID NOS: 850
```

```

: SOFTWARE: PERL Program
: SEQ ID NO 310
: LENGTH: 2794
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc feature
: OTHER INFORMATION: Incyte ID NO. US20020137081A1 060151CB1
US-10-044-090-330

Alignment Scores:
Pred. No.: 0.151 Length: 2794
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: 14 Gaps: 0

US-10-029-345A-109 (1-665) x US-10-044-090-330 (1-2794)
Oy 240 ValLeuValHisCybLeuAlaGlyTlLeserArgSer 251
|||
Db 1314 GTCTTGTAACATGCTTGCTGGCTGCATTAAGCGCTCA 1349

RESULT 38
US-10-199-221-12
: Sequence 12, Application US/10199221
: Publication No. US20040014048A1
: GENERAL INFORMATION:
: APPLICANT: Brett P. Monia
: APPLICANT: Lex M. Cowsett
: APPLICANT: Kenneth W. Dobie
: TITLE OF INVENTION: ANTISENSE MODULATION OF DUAL SPECIFIC PHOSPHATASE 6 EXPRESSION
: FILE REFERENCE: PTS-0009
: CURRENT APPLICATION NUMBER: US/10/199,221
: CURRENT FILING DATE: 2002-07-18
: NUMBER OF SEQ ID NOS: 101
: SEQ ID NO 12
: LENGTH: 7001
: TYPE: DNA
: ORGANISM: H. sapiens
: FEATURE:
US-10-199-221-12

Alignment Scores:
Pred. No.: 0.35 Length: 7001
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: 16 Gaps: 0

US-10-029-345A-109 (1-665) x US-10-199-221-12 (1-7001)
Oy 240 ValLeuValHisCybLeuAlaGlyTlLeserArgSer 251
|||
Db 4496 GTCTTGTAACATGCTTGCTGGCTGCATTAAGCGCTCA 4531

RESULT 39
US-10-085-117-241/c
: Sequence 241, Application US/10085117
: Publication No. US2003023334A1
: GENERAL INFORMATION:
: APPLICANT: Morris, David W.
: APPLICANT: Engelhard, Eric K.
: TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR CANCER
: FILE REFERENCE: 529452000121
: CURRENT APPLICATION NUMBER: US/10/085,117
: CURRENT FILING DATE: 2002-02-27
: PRIOR APPLICATION NUMBER: US 09/798,566
: PRIOR FILING DATE: 2001-03-02
: NUMBER OF SEQ ID NOS: 361
: SOFTWARE: FastSeq for Windows Version 4.0

```

```

; SEQ ID NO 241
; LENGTH: 165961
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: variation
; LOCATION: (1) ... (165961)
; OTHER INFORMATION: n = any nucleotide
US-10-085-117-241

Alignment Scores:
Pred. No.: 664 Length: 165961
Score: 10.00 Matches: 10
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.50% Indels: 0
DB: 16 Gaps: 0

US-10-029-345A-109 (1-665) x US-10-085-117-241 (1-165961)
Oy 300 AaNgInThrIglYAlAserGlyProLYser 309
|||||
Db 136778 AACCAACAGAGAGCAGCTGGCCCAAGAGC 136749
|||||

RESULT 40
US-10-006-285-423
; Sequence 423, Application US/10006285
; Publication No. US20030165854A1
; GENERAL INFORMATION:
; APPLICANT: Mary Jane Cunningham
; TITLE OF INVENTION: MARKER GENES RESPONDING TO TREATMENT WITH TOXINS
; FILE REFERENCE: PA-0039 US
; CURRENT APPLICATION NUMBER: US/10/006,285
; CURRENT FILING DATE: 2001-12-05
; NUMBER OF SEQ ID NOS: 514
; SOFTWARE: PERL Program
; SEQ ID NO 423
; LENGTH: 1571
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030165854A1 332413.6
US-10-006-285-423

Alignment Scores:
Pred. No.: 96.5 Length: 1571
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.35% Indels: 0
DB: 15 Gaps: 0

US-10-029-345A-109 (1-665) x US-10-006-285-423 (1-1571)
Oy 495 AlagInARgSerLeuLeuSerProLeu 503
|||||
Db 104 GCCCAGCGTTCCTCTCTCTCCACTT 130
|||||

RESULT 41
US-09-736-457-801
; Sequence 801, Application US/09736457
; Patent No. US20020168637A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darlick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane

```

```
/ APPLICANT: Fan, Liqun
/ APPLICANT: Wang, Aijun
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
/ FILE REFERENCE: 210121.478C15
/ CURRENT FILING DATE: 2000-12-13
/ NUMBER OF SEQ ID NOS: 1864
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 801
/ LENGTH: 1619
/ TYPE: DNA
/ ORGANISM: Homo sapien
US-09-736-457-801

Alignment Scores:
Pred. No.: 99.2          Length: 1619
Score: 9.00             Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.35%      Indels: 0
DB: 9                   Gaps: 0

US-10-029-345A-109 (1-665) x US-09-736-457-801 (1-1619)

QY 246 AlaglylleserAgsSerAlaThrIle 254
DB 569 GCGGGCATCTCGGGGTGCGCCACCATC 595

RESULT 42
US-09-902-941-801
/ Sequence 801, Application US/09902941
/ Patent No. US20020172952A1
/ GENERAL INFORMATION:
/ APPLICANT: Henderson, Robert A.
/ APPLICANT: Wang, Tonglong
/ APPLICANT: Watanabe, Yoshihiro
/ APPLICANT: Johnson, Jeffrey C.
/ APPLICANT: Retter, Marc W.
/ APPLICANT: Marnierakis, Margarita
/ APPLICANT: Carter, Darrick
/ APPLICANT: Fanger, Gary R.
/ APPLICANT: Vedvick, Thomas S.
/ APPLICANT: Bangur, Chaitanya S.
/ APPLICANT: McNabb, Andria
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
/ FILE REFERENCE: 210121.478C17
/ CURRENT FILING DATE: 2001-07-10
/ NUMBER OF SEQ ID NOS: 2002
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 801
/ LENGTH: 1619
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-902-941-801

Alignment Scores:
Pred. No.: 99.2          Length: 1619
Score: 9.00             Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.35%      Indels: 0
DB: 9                   Gaps: 0

US-10-029-345A-109 (1-665) x US-09-902-941-801 (1-1619)

QY 246 AlaglylleserAgsSerAlaThrIle 254
DB 569 GCGGGCATCTCGGGGTGCGCCACCATC 595

RESULT 43
```

```
US-09-849-626-801
/ Sequence 801, Application US/09849626
/ Publication No. US20020197669A1
/ GENERAL INFORMATION:
/ APPLICANT: Bangur, Chaitanya
/ APPLICANT: Fanger, Gary
/ APPLICANT: Wang, Aijun
/ APPLICANT: Wang, Tonglong
/ APPLICANT: Switzer, Anne
/ APPLICANT: McNeill, Patricia
/ APPLICANT: Clapper, Jonathan
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
/ FILE REFERENCE: 210121.478C16
/ CURRENT FILING DATE: 2001-05-03
/ NUMBER OF SEQ ID NOS: 1926
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 801
/ LENGTH: 1619
/ TYPE: DNA
/ ORGANISM: Homo sapien
US-09-849-626-801

Alignment Scores:
Pred. No.: 99.2          Length: 1619
Score: 9.00             Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.35%      Indels: 0
DB: 9                   Gaps: 0

US-10-029-345A-109 (1-665) x US-09-849-626-801 (1-1619)

QY 246 AlaglylleserAgsSerAlaThrIle 254
DB 569 GCGGGCATCTCGGGGTGCGCCACCATC 595

RESULT 44
US-10-283-017-801
/ Sequence 801, Application US/10283017
/ Publication No. US20030211510A1
/ GENERAL INFORMATION:
/ APPLICANT: Henderson, Robert A.
/ APPLICANT: Wang, Tonglong
/ APPLICANT: Watanabe, Yoshihiro
/ APPLICANT: Kalos, Michael D.
/ APPLICANT: Sleath, Paul R.
/ APPLICANT: Johnson, Jeffrey C.
/ APPLICANT: Retter, Marc W.
/ APPLICANT: Durham, Margarita
/ APPLICANT: Carter, Darrick
/ APPLICANT: Fanger, Gary R.
/ APPLICANT: Vedvick, Thomas S.
/ APPLICANT: Bangur, Chaitanya S.
/ APPLICANT: McNabb, Andria
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
/ FILE REFERENCE: 210121.478C20
/ CURRENT FILING DATE: 2002-10-28
/ NUMBER OF SEQ ID NOS: 2157
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 801
/ LENGTH: 1619
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-10-283-017-801

Alignment Scores:
Pred. No.: 99.2          Length: 1619
Score: 9.00             Matches: 9
Percent Similarity: 100.00% Conservative: 0
```


[illegible]

```

; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: PERL Program
; SEQ ID NO 51
; LENGTH: 1685
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Template ID: 235631.2
US-10-002-600-51

Alignment Scores:
Pred. No.: 103          Length: 1685
Score: 9.00           Matches: 9
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 1.35%          Indels: 0
DB: 14                 Gaps: 0

US-10-029-345A-109 (1-665) x US-10-002-600-51 (1-1685)

QY 246 AlaglylleSerArgSerAlaThrIle 254
DB 862 GCGGGTATCTCGCGCTCTGCCACCATC 888

RESULT 49
US-10-342-887-1034
; Sequence 1034, Application US/10342887
; Publication No. US20040058340A1
; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter S.
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Christopher J.
; APPLICANT: Van 't Veer, Laura Johanna
; APPLICANT: Van de Vijver, Marc J.
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-188-999
; CURRENT APPLICATION NUMBER: US/10/342,887
; CURRENT FILING DATE: 2003-01-15
; PRIOR APPLICATION NUMBER: 60/298,918
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: 60/380,710
; PRIOR FILING DATE: 2002-05-14
; PRIOR APPLICATION NUMBER: 10/172,118
; PRIOR FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 1034
; LENGTH: 1702
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-342-887-1034

Alignment Scores:
Pred. No.: 104          Length: 1702
Score: 9.00           Matches: 9
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 1.35%          Indels: 0
DB: 13                 Gaps: 0

US-10-029-345A-109 (1-665) x US-10-342-887-1034 (1-1702)

QY 246 AlaglylleSerArgSerAlaThrIle 254
DB 860 GCGGGTATCTCGCGCTCTGCCACCATC 886

RESULT 50
US-10-172-118-1034
; Sequence 1034, Application US/10172118
; Publication No. US20030224374A1
```

```

; GENERAL INFORMATION:
; APPLICANT: Dai, Hongyue
; APPLICANT: He, Yudong
; APPLICANT: Linsley, Peter
; APPLICANT: Mao, Mao
; APPLICANT: Roberts, Chris
; APPLICANT: Van 't Veer, Laura
; APPLICANT: Van de Vijver, Marc
; APPLICANT: Bernards, Rene
; TITLE OF INVENTION: Diagnosis and Prognosis of Breast Cancer Patients
; FILE REFERENCE: 9301-175-999
; CURRENT APPLICATION NUMBER: US/10/172,118
; CURRENT FILING DATE: 2002-06-14
; PRIOR APPLICATION NUMBER: 60/380,770
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 2699
; SEQ ID NO 1034
; LENGTH: 1702
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: NM_004418
; DATABASE ENTRY DATE: 2001-06-18
US-10-172-118-1034

Alignment Scores:
Pred. No.: 104          Length: 1702
Score: 9.00           Matches: 9
Percent Similarity: 100.00%  Conservative: 0
Best Local Similarity: 100.00%  Mismatches: 0
Query Match: 1.35%          Indels: 0
DB: 13                 Gaps: 0

US-10-029-345A-109 (1-665) x US-10-172-118-1034 (1-1702)

QY 246 AlaglylleSerArgSerAlaThrIle 254
DB 860 GCGGGTATCTCGCGCTCTGCCACCATC 886
```

Search completed: June 22, 2004, 10:25:07
Job time : 835 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: June 22, 2004, 06:30:22 ; Search time 122 Seconds

(without alignments)
3024.937 Million cell updates/sec

Title: US-10-029-345A-109

Perfect score: 665
Sequence: 1 MAHEMIGTIVTERLVALLE.....LKVSGSSFSGSMRIEVS 665

Scoring table:

OLIGO	Xgapop 60.0 , Xgapext 60.0
	Ygapop 60.0 , Ygapext 60.0
	Fgapop 6.0 , Fgapext 7.0
	Delop 6.0 , Delext 7.0

Searched: 682709 seqs, 277475446 residues

Word size: 1

Total number of hits satisfying chosen parameters: 92308

Minimum DB seq length: 1419

Maximum DB seq length: 200000000

Post-processing: Listing first 75 summaries

Command line parameters:

```
-MODE=frame+ p2n.model -DEV=x1h  
-CGN2_1/USPTO.spool/US10029345.rn1 21062004.164401.29776/APP.query.fasta_1.839  
-DB=Issued Patents NA -QFMT=fasta -SUFFIX=oligo.rn1 -MINMATCH=0.1 -LOOPEXT=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=75 -DCCALIGN=200 -THR_SCORE=quality -THR_MIN=1 -ALIGN=50 -MODE=LOCAL  
-OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=1419 -MAXLEN=200000000  
-USER=US10029345.CCN_1_1.56.rn1 21062004.164401.29776 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7
```

Database :

```
Issued Patents_NA:*  
1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*  
2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*  
3: /cgn2_6/prodata/2/ina/5A_COMB.seq:*  
4: /cgn2_6/prodata/2/ina/5B_COMB.seq:*  
5: /cgn2_6/prodata/2/ina/5A_COMB.seq:*  
6: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	472	71.0	1998	4 US-09-816-494-3	Sequence 3, Appl1
2	472	71.0	3544	4 US-09-816-494-1	Sequence 1, Appl1
3	20	3.0	2377	4 US-09-920-668-3	Sequence 1, Appl1
4	12	1.8	2109	4 US-09-016-434-1135	Sequence 1135, Ap
5	12	1.8	2109	4 US-09-023-655-946	Sequence 946, App
6	9	1.4	1619	4 US-09-702-705-801	Sequence 801, App
7	9	1.4	1619	4 US-09-736-457-801	Sequence 801, App
8	9	1.4	1619	4 US-09-614-124B-801	Sequence 801, App
9	9	1.4	1619	4 US-09-589-184-825	Sequence 801, App
10	9	1.4	1619	4 US-09-589-184-801	Sequence 801, App
11	9	1.4	1987	2 US-08-390-379-1	Sequence 1, Appl1
12	9	1.4	1993	2 US-08-390-379-2	Sequence 2, Appl1

13	9	1.4	2000	4 US-09-016-434-1191	Sequence 1291, Ap
14	9	1.4	2064	4 US-09-702-705-825	Sequence 825, App
15	9	1.4	2064	4 US-09-736-457-825	Sequence 825, App
16	9	1.4	2064	4 US-09-614-124B-825	Sequence 825, App
17	9	1.4	2064	4 US-09-671-325-825	Sequence 825, App
18	9	1.4	2064	4 US-09-589-184-825	Sequence 825, App
19	9	1.4	2109	4 US-09-702-705-826	Sequence 826, App
20	9	1.4	2109	4 US-09-736-457-826	Sequence 826, App
21	9	1.4	2109	4 US-09-614-124B-826	Sequence 826, App
22	9	1.4	2109	4 US-09-671-325-826	Sequence 826, App
23	9	1.4	2109	4 US-09-589-184-826	Sequence 826, App
24	9	1.4	2240	4 US-09-016-434-1100	Sequence 1100, Ap
25	9	1.4	4637	4 US-09-702-705-804	Sequence 804, App
26	9	1.4	4637	4 US-09-736-457-804	Sequence 804, App
27	9	1.4	4637	4 US-09-614-124B-804	Sequence 804, App
28	9	1.4	4637	4 US-09-671-325-804	Sequence 804, App
29	9	1.4	4637	4 US-09-589-184-804	Sequence 804, App
30	9	1.2	1421	5 PCT-US92-08090-3	Sequence 804, App
31	8	1.2	1426	2 US-08-284-465-2	Sequence 3, Appl1
32	8	1.2	1426	2 US-08-284-465-7	Sequence 2, Appl1
33	8	1.2	1426	2 PCT-US92-08090-2	Sequence 2, Appl1
34	8	1.2	1508	1 US-08-236-311-5	Sequence 5, Appl1
35	8	1.2	1508	1 US-08-457-918-5	Sequence 5, Appl1
36	8	1.2	1599	2 US-08-417-495-3	Sequence 3, Appl1
37	8	1.2	1599	2 US-08-284-391B-3	Sequence 3, Appl1
38	8	1.2	1599	3 US-09-218-950-3	Sequence 3, Appl1
39	8	1.2	1599	3 PCT-US92-01785-3	Sequence 3, Appl1
40	8	1.2	1599	5 PCT-US95-00454-3	Sequence 3, Appl1
41	8	1.2	1669	2 US-08-916-902A-2	Sequence 2, Appl1
42	8	1.2	1669	2 US-09-213-389-2	Sequence 2, Appl1
43	8	1.2	1728	2 US-08-417-495-1	Sequence 1, Appl1
44	8	1.2	1728	2 US-08-284-391B-1	Sequence 1, Appl1
45	8	1.2	1728	3 US-09-218-950-1	Sequence 1, Appl1
46	8	1.2	1728	5 PCT-US92-01785-1	Sequence 1, Appl1
47	8	1.2	1728	5 PCT-US95-00454-1	Sequence 1, Appl1
48	8	1.2	1742	4 US-08-466-368-3	Sequence 3, Appl1
49	8	1.2	1742	4 US-09-517-605-7	Sequence 7, Appl1
50	8	1.2	1742	4 US-08-470-999-1	Sequence 1, Appl1
51	8	1.2	1742	4 US-09-023-655-1013	Sequence 1013, Ap
52	8	1.2	1742	4 US-08-328-500-8	Sequence 8, Appl1
53	8	1.2	1796	3 US-08-477-460B-1	Sequence 1, Appl1
54	8	1.2	1796	3 US-08-379-516-1	Sequence 1, Appl1
55	8	1.2	1796	3 US-09-329-916-1	Sequence 1, Appl1
56	8	1.2	1796	3 US-08-485-372A-1	Sequence 1, Appl1
57	8	1.2	1796	4 US-09-409-006A-1	Sequence 1, Appl1
58	8	1.2	1796	4 US-08-484-681-1	Sequence 1, Appl1
59	8	1.2	1796	5 PCT-US93-07422-1	Sequence 1, Appl1
60	8	1.2	1858	3 US-08-813-574-1	Sequence 1, Appl1
61	8	1.2	1861	4 US-09-252-991A-9705	Sequence 9705, Ap
62	8	1.2	2303	4 US-09-922-146-3	Sequence 3, Appl1
63	8	1.2	2465	5 PCT-US92-08090-1	Sequence 3, Appl1
64	8	1.2	2482	3 US-08-477-460B-3	Sequence 3, Appl1
65	8	1.2	2482	3 US-08-379-516-3	Sequence 3, Appl1
66	8	1.2	2482	3 US-09-329-916-3	Sequence 3, Appl1
67	8	1.2	2482	3 US-08-485-372A-3	Sequence 3, Appl1
68	8	1.2	2482	4 US-09-409-006A-3	Sequence 3, Appl1
69	8	1.2	2482	4 US-08-484-681-3	Sequence 3, Appl1
70	8	1.2	2482	5 PCT-US93-07422-3	Sequence 3, Appl1
71	8	1.2	2589	4 US-08-472-888A-5	Sequence 5, Appl1
72	8	1.2	2709	1 US-08-021-601-11	Sequence 11, Appl1
73	8	1.2	2709	1 US-08-082-849B-11	Sequence 11, Appl1
74	8	1.2	2709	5 PCT-US94-01624-11	Sequence 11, Appl1
75	8	1.2	2940	2 US-08-418-848A-3	Sequence 3, Appl1

ALIGNMENTS

RESULT 1
US-09-816-494-3
; Sequence 3, Application US/09816494
; Patent No. 6664089
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel A.

; TITLE OF INVENTION: 38692 AND 21117, NOVEL DUAL SPECIFICITY
 ; TITLE OF INVENTION: PHOSPHATASE MOLECULES AND USES THEREFOR
 ; FILE REFERENCE: 10448-030002
 ; CURRENT APPLICATION NUMBER: US/09/816,494
 ; CURRENT FILING DATE: 2001-03-23
 ; PRIOR APPLICATION NUMBER: US 60/191,858
 ; PRIOR FILING DATE: 2000-03-24
 ; NUMBER OF SEQ ID NOS: 10
 ; SOFTWARE: FASTSEQ for Windows Version 4.0
 ; SEQ ID NO 3
 ; LENGTH: 1998
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-09-816-494-3

Alignment Scores:

Pred. No.:	0	Length:	1998
Score:	472.00	Matches:	663
Percent Similarity:	99.40%	Conservative:	0
Best Local Similarity:	99.40%	Mismatches:	2
Query Match:	70.98%	Indels:	4
DB:	4	Gaps:	0

US-10-029-345a-109 (1-665) x US-09-816-494-3 (1-1998)

```

QY      1 MetAlaHisGluMetCllEgLYThrGlnIleValThrGluArgLeuValAlaLeuLeuGlu
DB      1 ATGGCCCATGAGTGGATGGACATCCAAATGCTTACTGAGAGGTTGGGCTGCTGAGAA
QY      21 SerGlyThrGluLysValLeuLeuLleAspSerArgProPheValGluTyrAsnThrSer
DB      61 AGTGAACCGGAAAAAGGCTGCTAATGATGAGCGGCCCATTTGTGAATATCAATATCATCC
QY      41 HisIleLeuGluAlaIleAsnIleAsnCysSerLysLeuMetLysArgArgLeuGlnGln
DB      121 CACATTTTGGAAACCATTAATATCAACTGCTCCAGCTTATGAGGAGGATTGGCAACG
QY      61 AspLysValLeuLleThrGluLeuLleGlnHisSerLysLysHisLysValAspIleAsp
DB      181 GACAAAGTGTTAATTAACAGAGCTCATCCAGCATTCAGCGAAACATTAAGTTGACATTGAT
QY      81 CysSerGlnLysValValValTyrAspGlnSerSerGlnAspValAlaSerLeuSerSer
DB      241 TGAAGTCAGAAAGGTGTGATTACGATCAAAAGCTCCCAAGATGTGGCTCTCTCTTCA
QY      101 AspCysPheLeuThrValLeuLeuGlyLysLeuGluLysSerPheAsnSerValHisLeu
DB      301 GACTGTTTCTCATCTTACTTCTGGGTAACAGAGAGAGCTTCAACTGCTTCACTG
QY      121 LeuAlaGlyGlyPheAlaGluPheSerArgCysPheProGlyLeuCysGluGlyLysSer
DB      361 CTTCAGAGGTGGTGTGCTGAGTTCCTCGTGTTCCTCGGCTCTGAGAGGAAATCC
QY      141 ThrLeuValProThrCysIleSerGlnProCysLeuProValAlaAsnIleGlyProThr
DB      421 ACTCTAGTCCCTACCTGCACTTCTCAGCTTCCTTACCTGTTGCCAAATGGGCAACC
QY      161 ArgGlyLeuProAsnLeuTyrLeuGlyCysGlnArgAspValLeuAsnLysGlyLeuLle
DB      481 CCAATTCCTCCCAATCTTATCTTGCTGCTGCGCAGCGAGATGCTCTCAACAAGAGCTGAT-
QY      181 -GlnGlnAsnGlyIleGlyTyrValLeuAsnAlaSerTyr-ThrCysProLysProAsp
DB      540 GCAAGCAAGATGGAGATTGTTATGTGTAAATGCCAGCAA-TACCTGTCCAAAGCCTGACT
QY      200 heIleProGlnSerHisPheLeuArgValProValAsnAspSerPheCysGluLysIleL
DB      599 TTATCCCGAGAGCTCATTTCCGCGCTGTGCTGTGAAGACAGCTTTGTGAGAAATTT
QY      220 euProTyrLeuAspLysSerValAspPheIleGluLysAlaLysAlaSerAsnGlyCysV
DB      659 TGCCGTGTGTGACAAATCAGTAGATTCAATTGAGAAAGCAAAAGCCTCCCAATGGATGTG
  
```

```

QY      240 alLeuValHisCysLeuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIleM
DB      719 TTCTAGTGCACGTGTTAGTGGATCTCCCGCTCCGCCACCATGCTTACCTTACATCA
QY      260 eLysArgMetAspMetSerLeuAspGluAlaTyrArgPheValLysGluLysArgProT
DB      779 TGAAGAGATGAGCAATGCTTATGATGAAAGCTTACAGATTGTGTGAAGAAAAAGACCTA
QY      280 hrIleSerProAsnPheAsnPheLeuGlyLysIleLeuAspTyrGluLysLysIleLysA
DB      839 CTATATCTCCCAACTTCAATTTTCTGGGCCCACTCTGAACTATGAGAGAAAGATTAAGA
QY      300 SerGlnThrGlyAlaSerGlyProLysSerLysLysLeuLeuHisIleGluGluLysProA
DB      899 ACCAGACTGAGACATCAGGCGCAAGAGCAAACTCAAGCTGCTGCACTGGAAGAGCCAA
QY      320 SerGlnProValProAlaValSerGluGlyGlnLysSerGlnThrProLeuSerProp
DB      959 ATGAACCTGTCTCCCTGCTGTCTTCAGAGGAGTGAGCAAGAAAGGAGACCGCCCTCAGTCCAC
QY      340 roCysAlaAspSerAlaThrSerGluAlaAlaGlyGlnArgProValHisProAlaSerV
DB      1019 CCTGTGCGGACCTGCTGCTACTCAGAGGCGAGAGCAAAAGCCCTGCTGATCCGCCAGCG
QY      360 alProSerValProSerValGlnProSerLeuLeuGluAspSerProLeuValGlnAlaL
DB      1079 TGCCAGCGGTGCCAGCGCTGCGAGCGCTGCTGTATGAGAGACAGCCCGCTGTACAGCGC
QY      380 euSerGlyLysHisIleLeuSerAlaAspArgLeuGluAspSerAsnLysLysArgSerP
DB      1139 TCAGTGGGCTGCACCTGCTCCGACAGACAGCTGGAAGACAGCAATTAAGAGGTTCTT
QY      400 heSerLeuAspLysLysSerValSerTyrSerAlaSerMetAlaAlaSerLeuHisGlyP
DB      1199 TCTCTGAGATATCAATACATTTCATATTCAGGCAGACATGGACATCTTTCATATGCGCT
QY      420 heSerSerSerGluAspAlaLeuGluLysTyrLysProSerThrThrLeuAspGlyThra
DB      1259 TCTCTCATCAAGAAAGATCTTGTGAATATCAAAACCTTCACATCTCTGAGAGGAGCCA
QY      440 snLysLeuCysGlnPheSerProValGlnLysLeuSerGlnGlnThrProGluThSerP
DB      1319 ACAAGTATGCCAGGTTCTCCCGTTCAGGAACTATGAGAGCACTCCGAAACCAAGTTC
QY      460 roAspLysGluGluAlaSerIleProLysLysLeuGlnThrAlaArgProSerAspSerG
DB      1379 CTGATTAAGAGAGAGCAGCATCCCAAGAGCTGCAGACCGCCAGGCTTCAGACAGCC
QY      480 InSerLysArgLeuHisSerValArgThrSerSerSerGlyThrAlaGlnArgSerLeuL
DB      1439 AGAGCAAGCATTTGCATTCGGTCAAGAACAGAGAGTGCACCGCCAGAGGTCCTTTT
QY      500 euSerProLeuHisArgSerGlySerValGluAspAsnTyrHisThrSerPheLeuPheG
DB      1499 TATCTCAGCTGATCGAAGAGGAGGAGGCTGAGAGCAATTATACCAACCACTTCCTTTG
QY      520 LysLeuSerThrSerGlnGlnHisIleLeuThrLysSerAlaGlyLysLeuLysGlyTyrP
DB      1559 GCCCTTCCACAGACCGCAGAGCACTCAAGAAATGTGCTGCGGCTTTAAGGGCTGAGC
QY      540 isSerAspIleLeuAlaProGlnThrSerThrProSerLeuThrSerSerTyrPhePhA
DB      1619 ACTCGATATCTTGGGCCCCCAGAGACTTCAACCTTCCCTTGACAGAGAGCTGATTTTGG
QY      560 LathGlnSerSerHisPheTyrSerAlaSerAlaIleTyrGlyLysSerAlaSerTyrS
DB      1679 CCAAGAGTCTTCACACTTCTACTGCTGCTGAGCATCATCAGAGAGAGAGTGCAGATTACT
QY      580 eAlaIleTyrSerCysSerGlnLysProThrCysGlyAspGlnValTyrSerValArgArgA
DB      1739 CTGCTTACAGCTGAGCGCAGCTGCGCACTTGCGGAGAGCAACATCTATTTCTGTGCGCAGCG
QY      600 rgGlnLysProSerAspArgAlaAspSerArgArgSerTyrHisGlnGluLysSerPropheG
  
```

|||||
Db 1799 GCGAGAGCCAGTGCACAGCTGACTCCGCGCGAGCTGGCATGAAGAGCCCTTTG 1858
|||
Qy 620 TULYSGINPHELYSARGARGSERCYSGIMMERCILUPHEGLYGIUSERIIMERSERGLN 640
|||
Db 1859 AAAAGCAGTTTAAACCCAGAGCTGCCAATGTGAATTTGGAGAGACATCATGTACAGCA 1918
|||
Qy 640 GAARGSERARGIUGIULYSGIYLYSVALGYSERGINSESRPHESESGLYSERMERC 660
|||
Db 1919 ACAGGTCCAGCGAGAGCTGGGAGAAAGTGGCAGTCACTTTCGGGCGAGCATG 1978
|||
Qy 660 TULIEGLEUVALSER 665
|||
Db 1979 AATTCATTGAGCTCTCC 1995
|||
RESULT 2
US-09-816-494-1
; Sequence 1, Application US/09816494
; Patent No. 6664089
; GENERAL INFORMATION:
; APPLICANT: Meyers, Rachel A.
; TITLE OF INVENTION: 38692 AND 21117, NOVEL DUAL SPECIFICITY
; TITLE OF INVENTION: PHOSPHATASE MOLECULES AND USES THEREFOR
; FILE REFERENCE: 10448-030002
; CURRENT APPLICATION NUMBER: US/09/816,494
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 60/191,858
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 3544
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (589)...(2583)
US-09-816-494-1

Alignment Scores:
Pred. No.: 0 Length: 3544
Score: 472.00 Matches: 663
Percent Similarity: 99.40% Conservative: 0
Best Local Similarity: 99.40% Mismatches: 2
Query Match: 70.98% Indels: 4
DB: 4 Gaps: 0

US-10-029-345a-109 (1-665) x US-09-816-494-1 (1-3544)
Qy 1 MetAlHhIGIuMeCIEGLYThrGlnIleValThrGlnArgLeuValAlaLeuLeuGln 20
|||
Db 589 ATGGCCCATGAGATGATTTGAGACTCAAAATGTACTGAGAGTGGCTCTGCTGGAA 648
|||
Qy 21 SerGILThrGlnValLeuLeuIleAspSerArgProPheValGluTyrAsnThrSer 40
|||
Db 649 AGTGAACGGAAGAAAGTGTCTTAATGATGACCGGCATTGTGGGATACAAATACATCC 708
|||
Qy 41 HisIleLeuGlnAlaIleAsnIleAsnCySerIleLeuMetIleYsArgArgLeuGln 60
|||
Db 709 CACATTTTGGAAACCATTAATATCACTGCTCCAGCTTATGAAGCAAGATTGCCAAG 768
|||
Qy 61 AspIleValIleIleThrGlnLeuIleGlnHisSerAlaIleYsIleValAspIleAsp 80
|||
Db 769 GACAAAGTGTAAATTCAGAGCTCATCTCCAGCATTCAGCAACATTAAGTTGACATTTGAT 828
|||
Qy 81 CysSerGlnIleValIleValTyrAspGlnSerSerGlnAspValAlaSerIleAspSer 100
|||
Db 829 TGCAGTCAGAAAGTTGATTTACGATCAAGCTCCCAAGATGTTGCTCTCTCTTCA 888
|||
Qy 101 AspCyPheLeuThrValIleLeuGlnIleValLeuGlnIleYsSerPheAsnSerValIleLeu 120
|||
Db 889 GACGTGTTTCTCACTGTACTTCTGGGTAACTGAGAGAGCTTCAACTCTGTTCACTG 948
|||

Qy 121 LeuAlaGlyIlePheAlaGluPheSerArgCysPheProGlyLeuCySGluGlyIleYsSer 140
|||
Db 949 CTTCGAGGTGGGTTTCTGAGTTCTCTGCTGTGTTCCTGGCTCTGTGTAGAGAAATTC 1008
|||
Qy 141 ThrIleValProThrCysIleSerGlnProCysLeuProValAlaAsnIleGlyProThr 160
|||
Db 1009 ACTCTAGTCCCTACCTGACATTTCTCAGCTTCTTCACTGTTCAGCAACATTTGGGCCAAC 1068
|||
Qy 161 ArgIleLeuProAsnLeuTyrIleuGlyCysGlnArgAspValLeuAsnIleGluLeuIle 180
|||
Db 1069 CGAATTTCTCCAAATCTTATCTGTGGCTCCAGCGAGATGCTCTCAACAGAGCTGAT- 1127
|||
Qy 181 -GlnGlnAsnGlyIleGlyTyrValIleuAsnAlaSerTyr-ThrCysProIleProAsp 200
|||
Db 1128 GACGACGAATGGGATTTGTTATGTGTTAAATGCCAGAA-TACTGTCCAAAGCTGACT 1186
|||
Qy 200 IleIleProGluSerHisPheLeuArgValProValAsnAspSerPheCysGluYsIle 220
|||
Db 1187 TTATCCCGAGTCTGATTTCTGCGGTGCTGTGATGACAGCTTTGTGAGAAATTT 1246
|||
Qy 220 euProTTrLeuAspIleYsSerValAspPheIleGluValAlaIleAsnAsnGlyCys 240
|||
Db 1247 TGCCTGCTTGGACAAATCAGATGATTTCAATTGAGAAACAAAGCCTCAATGATGTG 1306
|||
Qy 240 AlIleValHisCysIleuAlaGlyIleSerArgSerAlaThrIleAlaIleAlaTyrIle 260
|||
Db 1307 TTTCTAGTCACTGTTAGCTGGGATCTCCGCTCCGACCAATGGTATGCTTACATCA 1366
|||
Qy 260 eLlyeArgMetAspMetSerLeuAspGluAlaTyrArgPheValYsGluYsArgProT 280
|||
Db 1367 TGAAGAGATGGACATGCTTTAGATGAGAGCTTACAGATTTGGAAAGAAAGACCTTA 1426
|||
Qy 280 hrIleSerProAsnPheAsnPheLeuGlyGlnLeuLeuAspTyrGluYsIleYs 300
|||
Db 1427 CTATATCTCCAACTTCAATTTTCTGGGCCAATCTCTGACATATGAGAAAGATTAAAG 1486
|||
Qy 300 GlnGlnThrGlyIleAsnGlyProIleYsSerIleuYsLeuIleuHisIleGluIleYsPro 320
|||
Db 1487 ACCAGACTGAGATCAGGAGCCAAAGAGCAAACTCAAGCTGTGCACTGGAGAACCA 1546
|||
Qy 320 GlnIleProValProAlaValSerGluGlyGlyIleYsSerGluThrProLeuSerPro 340
|||
Db 1547 ATGAACCTTCTCCTGCTGTCTGACAGAGGTGGACAGAAAGCCAGAGCCCTCAGTCC 1606
|||
Qy 340 roCyAlaAspSerAlaThrSerGluAlaIleGlyGlnArgProValHisProAlaSer 360
|||
Db 1607 CCGTGCACACTGCTGCTACCTCAGAGGACAGCAAGAGCCCGTGCATCCGCCAGCG 1666
|||
Qy 360 AlProSerValProSerValGlnProSerIleuGlnIleAspSerProLeuValGlnAla 380
|||
Db 1667 TGCCCGCGTCCGACGCTGACGCTGCTGTGATGAGACAGCCGCTGTGACAGGCG 1726
|||
Qy 380 euSerGlyLeuHisLeuSerAlaAspArgLeuGluAspSerAsnIleYsLeuYsArgSer 400
|||
Db 1727 TCAAGTGGCTGACACTGCTCCGACAGACAGCTGGAACAGCAATAGCTCAAGCTTCT 1786
|||
Qy 400 HeSerLeuAspIleYsSerValSerTyrSerAlaSerMetAlaIleSerLeuHisGlyP 420
|||
Db 1787 TCTCTCTGATATCAATACAGTTTCATATTACGCCAGCTAGGACAGATCTTCAATGGCT 1846
|||
Qy 420 HeSerSerSerGluAspAlaLeuGluTyrTyrIleYsProSerThrThrLeuAspGlyThr 440
|||
Db 1847 TCTCTCATCAGAAAGTGTGTTGGAATCTACAAACCTTCACTACTGATGGAGCA 1906
|||
Qy 440 snIleLeuCyGlnPheSerProValGlnIleuSerGluGlnThrProGluThrSerP 460
|||
Db 1907 ACAAGCTAGCCAGATTCTCCCTGTTCAAGAACTATCGAGAGCATCCCGAAACAGCT 1966
|||
Qy 460 roAspIleGluGlnAlaSerIleProYsIleLeuGlnIleThrAlaArgProSerAspSer 480
|||
Db 1967 CTGATTAAGAGAGAGCCACATCTCCAAAGAGCTGACAGACCGCCAGAGCTTCAAGAGCC 2026
|||
Qy 480 InSerIleYsArgLeuHisSerValArgThrSerSerSerGlyThrAlaGlnArgSerLeu 500
|||

DB 2027 AGAGGAGGATTGATTGCTGCTAGAACGACGAGCTGGACCGCCGAGAGGTCCCTTT 2086
QY 500 eUSErProleuHISaHsGserGlySerValGluHspAserYrHISrHsPheLeuPhg 520
DB 2087 TATCTCCACTGATCCAGAGTGGAGGCTGGAGACATTATCCACACACTTCTCTTTCG 2146
QY 520 lYleuSerHsSerGlnGlnHISleuThylSerValaglyleuGlyleuYrGlyTyrH 540
DB 2147 GCTTTTCCACCGACGACGACCTTACAGAGTCTCTGCTGCTGGCTTAAAGGCTTGGC 2206
QY 540 iSserAspIleleuAlaProGlnHISerThHsPheLeuHsSerSerTyrPhea 560
DB 2207 ACTGGATATCTTGGCCCCCAGACCTTACCCCTTCCCTGACAGACAGCTGGATTTCG 2266
QY 560 lATHrGluSerHsHISpHeYrYrSerAlaSerAlaIleYrGlyGlySerAlaSerTyrS 580
DB 2267 CCACAGAGTCTTCCACTTCTACTCTGCTCAGCCATCTTCAAGAGGCACTGCACTTACT 2326
QY 580 eRAlATySerCySserGlnLeuProThCyGglYaspGlnValYrSerValArgArg 600
DB 2327 CTGCTTACAGCTGACCGACGCTGCCACTTGGGAGACCAAGTCTATTCTTGGCGAGGC 2386
QY 600 rGglYlYsPProSerAspArgAlaAspSerArgArgSerTyrHISglYlYserProPhg 620
DB 2387 GCGAGAGGCAAGTGCAGAGCTGACTCGCGGCGAGCTGGCATGAGAGAGCCCTTTG 2446
QY 620 lYlYsGlnHelysAgaYrSerCyGglYmerGluHelyglYlYserIleMerSerGlu 640
DB 2447 AAAGAGAGTTTAAACGACAGAGCTGCAATGGAATTTGAGAGACATCATGTCAGAGA 2506
QY 640 sHArgSerArgglYlYleuGlylYsValGlySerGlnSerSerPheSerGlySerMerG 660
DB 2507 AAGGTACAGGAGAGGCTGGGAAAGTGGCAGTCACTTCTTTCGGGCAAGCATGG 2566
QY 660 lYlYleGluYlYser 665
DB 2567 AAATCATGTAGGTCTCC 2583
RESULT 3
US-09-920-668-3
/ Sequence 3, Application US/0920668
/ Patent No. 6482644
/ GENERAL INFORMATION:
/ APPLICANT: Lex M. Cowser
/ APPLICANT: Brett P. Monia
/ TITLE OF INVENTION: ANTISENSE MODULATION OF DUAL SPECIFIC PHOSPHATASE 8 EXPRESSION
/ FILE REFERENCE: RTS-0246
/ CURRENT FILING DATE: 2001-08-01
/ NUMBER OF SEQ ID NOS: 49
/ SEQ ID NO 3
/ LENGTH: 2377
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (135) .. (2012)
US-09-920-668-3
Alignment Scores:
Pred. No.: 1,55e-10 Length: 2377
Score: 20.00 Matches: 20
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 3.01% Indels: 0
DB: 4 Gaps: 0
US-10-029-345A-109 (1-665) x US-09-920-668-3 (1-2377)
QY 242 ValHISCySleuAlaGlylYleSerArgSerAlaHISleAlaIleAlaTyrIleMerYls 261
DB 864 GTCCACTGTGCTGGGCTATCTCCGCTCTGCGACCATGCGCATGCGCTTACATGAAG 923

RESULT 4
US-09-016-434-1135
/ Sequence 1135, Application US/09016434
/ Patent No. 6500938
/ GENERAL INFORMATION:
/ APPLICANT: Janice Au-Young
/ APPLICANT: Jeffrey J. Seilhamer
/ TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
/ TITLE OF INVENTION: PATHWAY GENE EXPRESSION
/ NUMBER OF SEQUENCES: 1490
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
/ STREET: 3174 PORTER DRIVE
/ CITY: PALO ALTO
/ STATE: CALIFORNIA
/ COUNTRY: USA
/ ZIP: 94304
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/016,434
/ FILING DATE: HERewith
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER:
/ FILING DATE:
/ CLASSIFICATION:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Zeller, Karen J.
/ REGISTRATION NUMBER: 37,071
/ REFERENCE/DOCKET NUMBER: PA-0002 US
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (650) 855-0555
/ TELEFAX: (650) 845-4166
/ INFORMATION FOR SEQ ID NO: 1135:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2109 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ IMMEDIATE SOURCE:
/ LIBRARY: GENBANK
/ CLONE: g1418933
US-09-016-434-1135
Alignment Scores:
Pred. No.: 0.0245 Length: 2109
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: 4 Gaps: 0
US-10-029-345A-109 (1-665) x US-09-016-434-1135 (1-2109)
QY 240 ValHISCySleuAlaGlylYleSerArgSer 251
DB 1216 GCTTGTGATCATGTGCTGGCTGGATTAAGCGCTCA 1251
RESULT 5
US-09-023-655-946
/ Sequence 946, Application US/09023655
/ Patent No. 6607879
/ GENERAL INFORMATION:
/ APPLICANT: Cocks, Benjamin G.
/ APPLICANT: Susan G. Stuart
/ APPLICANT: Jeffrey J. Seilhamer
/ TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
/ TITLE OF INVENTION: EXPRESSION

```

; NUMBER OF SEQUENCES: 1508
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,655
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0001 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 946:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2109 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENBANK
; CLONE: G1418933
; US-09-023-655-946

Alignment Scores:
Pred. No.: 0.0245 Length: 2109
Score: 12.00 Matches: 12
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.80% Indels: 0
DB: 4 Gaps: 0

US-10-029-345A-109 (1-665) x US-09-023-655-946 (1-2109)

QY 240 ValLeuValHicCySeLeuAlaGlyIleSerArgSer 251
Db 1216 GTCTGTGATACATGCTTGCTGGCATTAAGCCGCTCA 1251

RESULT 6
US-09-702-705-801
; Sequence 801, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:
; APPLICANT: Wang, Tonglong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C14
; CURRENT APPLICATION NUMBER: US/09/702,705
; CURRENT FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 1833

```

```

; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 801
; LENGTH: 1619
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-702-705-801

Alignment Scores:
Pred. No.: 23.3 Length: 1619
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.35% Indels: 0
DB: 4 Gaps: 0

US-10-029-345A-109 (1-665) x US-09-702-705-801 (1-1619)

QY 246 AlaGlyIleSerArgSerAlaThrIle 254
Db 569 GCGGCGATCTCGCGGTGCGGCACACATC 595

RESULT 7
US-09-736-457-801
; Sequence 801, Application US/09736457
; Patent No. 6509448
; GENERAL INFORMATION:
; APPLICANT: Wang, Tonglong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; APPLICANT: Wang, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736,457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 801
; LENGTH: 1619
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-736-457-801

Alignment Scores:
Pred. No.: 23.3 Length: 1619
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.35% Indels: 0
DB: 4 Gaps: 0

US-10-029-345A-109 (1-665) x US-09-736-457-801 (1-1619)

QY 246 AlaGlyIleSerArgSerAlaThrIle 254
Db 569 GCGGCGATCTCGCGGTGCGGCACACATC 595

RESULT 8
US-09-614-124B-801
; Sequence 801, Application US/09614124B
; Patent No. 6630574
; GENERAL INFORMATION:
; APPLICANT: Wang, Tonglong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom

```



```
/ APPLICANT: Carter, Darriek
/ APPLICANT: Retter, Marc
/ APPLICANT: Mannion, Jane
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
/ FILE REFERENCE: 210121.478C9
/ CURRENT APPLICATION NUMBER: US/09/614.124B
/ CURRENT FILING DATE: 2001-07-11
/ NUMBER OF SEQ ID NOS: 1668
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 801
/ LENGTH: 1619
/ TYPE: DNA
/ ORGANISM: Homo sapien
US-09-614-124B-801

Alignment Scores:
Pred. No.: 23.3          Length: 1619
Score: 9.00             Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.35%      Indels: 0
DB: 4                   Gaps: 0

US-10-029-345A-109 (1-665) x US-09-614-124B-801 (1-1619)

QY 246 AlaGlyIleSerArgSerAlaThrIle 254
DB 569 GCGGCATCTCGCGGTGCGCCACCATC 595

RESULT 9
US-09-671-325-801
/ Sequence 801, Application US/09671325
/ Patent No. 6667154
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Tonglong
/ APPLICANT: Bangur, Chaitanya S.
/ APPLICANT: Lodes, Michael A.
/ APPLICANT: Fanger, Gary
/ APPLICANT: Vedvick, Tom
/ APPLICANT: Carter, Darriek
/ APPLICANT: Retter, Marc
/ APPLICANT: Mannion, Jane
/ APPLICANT: Pan, Liqun
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
/ FILE REFERENCE: 210121.478C12
/ CURRENT APPLICATION NUMBER: US/09/671,325
/ CURRENT FILING DATE: 2000-09-26
/ NUMBER OF SEQ ID NOS: 1825
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 801
/ LENGTH: 1619
/ TYPE: DNA
/ ORGANISM: Homo sapien
US-09-671-325-801

Alignment Scores:
Pred. No.: 23.3          Length: 1619
Score: 9.00             Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.35%      Indels: 0
DB: 4                   Gaps: 0

US-10-029-345A-109 (1-665) x US-09-671-325-801 (1-1619)

QY 246 AlaGlyIleSerArgSerAlaThrIle 254
DB 569 GCGGCATCTCGCGGTGCGCCACCATC 595

RESULT 10
US-09-589-184-801

/ Sequence 801, Application US/09589184
/ Patent No. 666447
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Tonglong
/ APPLICANT: Bangur, Chaitanya S.
/ APPLICANT: Lodes, Michael A.
/ APPLICANT: Fanger, Gary
/ APPLICANT: Vedvick, Tom
/ APPLICANT: Carter, Darriek
/ APPLICANT: Retter, Marc
/ APPLICANT: Mannion, Jane
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
/ FILE REFERENCE: 210121.478C8
/ CURRENT APPLICATION NUMBER: US/09/589,184
/ CURRENT FILING DATE: 2000-06-05
/ NUMBER OF SEQ ID NOS: 827
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 801
/ LENGTH: 1619
/ TYPE: DNA
/ ORGANISM: Homo sapien
US-09-589-184-801

Alignment Scores:
Pred. No.: 23.3          Length: 1619
Score: 9.00             Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.35%      Indels: 0
DB: 4                   Gaps: 0

US-10-029-345A-109 (1-665) x US-09-589-184-801 (1-1619)

QY 246 AlaGlyIleSerArgSerAlaThrIle 254
DB 569 GCGGCATCTCGCGGTGCGCCACCATC 595

RESULT 11
US-08-990-379-1
/ Sequence 1, Application US/08990379
/ Patent No. 5998188
/ GENERAL INFORMATION:
/ APPLICANT: Stock, Philip J
/ APPLICANT: Mitera-Press, Anita
/ TITLE OF INVENTION: Mitogen Activated Protein Kinase Phosphatase cDNAs and
/ FILE REFERENCE: 4104-000322USA
/ CURRENT APPLICATION NUMBER: US/08/990,379
/ CURRENT FILING DATE: 1997-12-15
/ EARLIER APPLICATION NUMBER: PCT/US96/10402
/ EARLIER FILING DATE: 1996-06-14
/ EARLIER APPLICATION NUMBER: 60/000,263
/ EARLIER FILING DATE: 1995-06-16
/ NUMBER OF SEQ ID NOS: 19
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 1
/ LENGTH: 1987
/ TYPE: DNA
/ ORGANISM: Rattus norvegicus
US-08-990-379-1

Alignment Scores:
Pred. No.: 28.6          Length: 1987
Score: 9.00             Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.35%      Indels: 0
DB: 2                   Gaps: 0

US-10-029-345A-109 (1-665) x US-08-990-379-1 (1-1987)

QY 246 AlaGlyIleSerArgSerAlaThrIle 254
```

Db 1218 GCCGCATCTCTAGATCAGCCACCATC 1244
RESULT 12
US-08-990-379-2
Sequence 2, Application US/08990379
Patent No. 5998188
GENERAL INFORMATION:
APPLICANT: Miera, Phillip J
TITLE OF INVENTION: Mitogen Activated Protein Kinase Phosphatase cDNAs and
FILE REFERENCE: 4104-000322USA
CURRENT APPLICATION NUMBER: US/08/990,379
EARLIER FILING DATE: 1997-12-15
EARLIER APPLICATION NUMBER: PCT/US96/10402
EARLIER FILING DATE: 1996-06-14
EARLIER APPLICATION NUMBER: 60/000,263
NUMBER OF SEQ ID NOS: 19
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 2
LENGTH: 1993
TYPE: DNA
ORGANISM: Rattus norvegicus
US-08-990-379-2
Alignment Scores:
Pred. No.: 28.7 Length: 1993
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.35% Indels: 0
DB: 2 Gaps: 0
US-10-029-345A-109 (1-665) x US-08-990-379-2 (1-1993)
QY 246 AlaGlyIleSerArgSerAlaThrIle 254
Db 1224 GCCGCATCTCTAGATCAGCCACCATC 1250
RESULT 13
US-09-016-434-1291
Sequence 1291, Application US/09016434
Patent No. 6500938
GENERAL INFORMATION:
APPLICANT: Janice Au-Young
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
FILE REFERENCE: 210121.478C14
CURRENT APPLICATION NUMBER: US/09/016,434
NUMBER OF SEQUENCES: 1490
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/016,434
FILING DATE: HERewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Zeller, Karen J.

REGISTRATION NUMBER: 37,071
REFERENCE/DOCKET NUMBER: PA-0002 US
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 1291:
SEQUENCE CHARACTERISTICS:
LENGTH: 2000 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GENBANK
CLONE: G29980
US-09-016-434-1291
Alignment Scores:
Pred. No.: 28.8 Length: 2000
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.35% Indels: 0
DB: 4 Gaps: 0
US-10-029-345A-109 (1-665) x US-09-016-434-1291 (1-2000)
QY 246 AlaGlyIleSerArgSerAlaThrIle 254
Db 1011 GCAGCATTTCCGCTCAGCCACCATC 1037
RESULT 14
US-09-702-705-825
Sequence 825, Application US/09702705
Patent No. 6504010
GENERAL INFORMATION:
APPLICANT: Wang, Tonglong
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Lodges, Michael A.
APPLICANT: Fanger, Gary
APPLICANT: Vedvick, Tom
APPLICANT: Carter, Darick
APPLICANT: Retter, Marc
APPLICANT: Mannion, Jane
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.478C14
CURRENT APPLICATION NUMBER: US/09/702,705
CURRENT FILING DATE: 2000-10-30
NUMBER OF SEQ ID NOS: 1833
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 825
LENGTH: 2064
TYPE: DNA
ORGANISM: Homo sapiens
US-09-702-705-825
Alignment Scores:
Pred. No.: 29.7 Length: 2064
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.35% Indels: 0
DB: 4 Gaps: 0
US-10-029-345A-109 (1-665) x US-09-702-705-825 (1-2064)
QY 246 AlaGlyIleSerArgSerAlaThrIle 254
Db 1240 GCCGCATCTCTAGATCAGCCACCATC 1266
RESULT 15
US-09-736-457-825

```
/ Sequence 825, Application US/09736457
/ Patent No. 6509448
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Tongtong
/ APPLICANT: Bangur, Chaitanya S.
/ APPLICANT: Lodes, Michael A.
/ APPLICANT: Fanger, Gary
/ APPLICANT: Vedvick, Tom
/ APPLICANT: Carter, Darrick
/ APPLICANT: Retter, Marc
/ APPLICANT: Mannion, Jane
/ APPLICANT: Fan, Liqun
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
/ FILE REFERENCE: 210121.478C15
/ CURRENT APPLICATION NUMBER: US/09/736,457
/ CURRENT FILING DATE: 2000-12-13
/ NUMBER OF SEQ ID NOS: 1864
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 825
/ LENGTH: 2064
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-736-457-825

Alignment Scores:
Pred. No.: 29.7 Length: 2064
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.35% Indels: 0
DB: 4 Gaps: 0

US-10-029-345A-109 (1-665) x US-09-736-457-825 (1-2064)
OY 246 AlAGlylleSerArgSerAlaThrIle 254
Db 1240 GCGGGCATCTCGCGTGGCCACCATC 1266

RESULT 16
US-09-614-124B-825
/ Sequence 825, Application US/09614124B
/ Patent No. 6630574
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Tongtong
/ APPLICANT: Bangur, Chaitanya S.
/ APPLICANT: Lodes, Michael A.
/ APPLICANT: Fanger, Gary
/ APPLICANT: Vedvick, Tom
/ APPLICANT: Carter, Darrick
/ APPLICANT: Retter, Marc
/ APPLICANT: Mannion, Jane
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
/ FILE REFERENCE: 210121.478C9
/ CURRENT APPLICATION NUMBER: US/09/614,124B
/ CURRENT FILING DATE: 2001-07-11
/ NUMBER OF SEQ ID NOS: 1668
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 825
/ LENGTH: 2064
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-614-124B-825

Alignment Scores:
Pred. No.: 29.7 Length: 2064
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.35% Indels: 0
DB: 4 Gaps: 0
```

```
US-10-029-345A-109 (1-665) x US-09-614-124B-825 (1-2064)
OY 246 AlAGlylleSerArgSerAlaThrIle 254
Db 1240 GCGGGCATCTCGCGTGGCCACCATC 1266

RESULT 17
US-09-671-325-825
/ Sequence 825, Application US/09671325
/ Patent No. 6667154
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Tongtong
/ APPLICANT: Bangur, Chaitanya S.
/ APPLICANT: Lodes, Michael A.
/ APPLICANT: Fanger, Gary
/ APPLICANT: Vedvick, Tom
/ APPLICANT: Carter, Darrick
/ APPLICANT: Retter, Marc
/ APPLICANT: Mannion, Jane
/ APPLICANT: Fan, Liqun
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
/ FILE REFERENCE: 210121.478C12
/ CURRENT APPLICATION NUMBER: US/09/671,325
/ CURRENT FILING DATE: 2000-09-26
/ NUMBER OF SEQ ID NOS: 1825
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 825
/ LENGTH: 2064
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-671-325-825

Alignment Scores:
Pred. No.: 29.7 Length: 2064
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.35% Indels: 0
DB: 4 Gaps: 0

US-10-029-345A-109 (1-665) x US-09-671-325-825 (1-2064)
OY 246 AlAGlylleSerArgSerAlaThrIle 254
Db 1240 GCGGGCATCTCGCGTGGCCACCATC 1266

RESULT 18
US-09-589-184-825
/ Sequence 825, Application US/09589184
/ Patent No. 6686447
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Tongtong
/ APPLICANT: Bangur, Chaitanya S.
/ APPLICANT: Lodes, Michael A.
/ APPLICANT: Fanger, Gary
/ APPLICANT: Vedvick, Tom
/ APPLICANT: Carter, Darrick
/ APPLICANT: Retter, Marc
/ APPLICANT: Mannion, Jane
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
/ FILE REFERENCE: 210121.478C8
/ CURRENT APPLICATION NUMBER: US/09/589,184
/ CURRENT FILING DATE: 2000-06-05
/ NUMBER OF SEQ ID NOS: 827
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO 825
/ LENGTH: 2064
/ TYPE: DNA
/ ORGANISM: Homo sapiens
US-09-589-184-825
```

Alignment Scores:

Pred. No.:	29.7	Length:	2064
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.35%	Indels:	0
DB:	4	Gaps:	0

US-10-029-345A-109 (1-665) x US-09-589-184-825 (1-2064)

QY 246 A1AG1Y1leSerArgerAlathr1le 254
DB 1240 GCGGCATCTCGCGGTCCGCCACCATC 1266

RESULT 19

US-09-702-705-826
; Sequence 826, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:
; APPLICANT: Wang, Tonglong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.478C14
; CURRENT APPLICATION NUMBER: US/09/702.705
; CURRENT FILING DATE: 2000-10-30
; NUMBER OF SEQ ID NOS: 1833
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 826
; LENGTH: 2109
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-702-705-826

Alignment Scores:

Pred. No.:	30.4	Length:	2109
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.35%	Indels:	0
DB:	4	Gaps:	0

US-10-029-345A-109 (1-665) x US-09-702-705-826 (1-2109)

QY 246 A1AG1Y1leSerArgerAlathr1le 254
DB 1200 GCGGCATCTCGCGGTCCGCCACCATC 1226

RESULT 20

US-09-736-457-826
; Sequence 826, Application US/09736457
; Patent No. 6509448
; GENERAL INFORMATION:
; APPLICANT: Wang, Tonglong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; APPLICANT: Fan, Liqun
; APPLICANT: Wang, Aijun
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER

FILE REFERENCE: 210121.478C15
; CURRENT APPLICATION NUMBER: US/09/736,457
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 1864
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 826
; LENGTH: 2109
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-736-457-826

Alignment Scores:

Pred. No.:	30.4	Length:	2109
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.35%	Indels:	0
DB:	4	Gaps:	0

US-10-029-345A-109 (1-665) x US-09-736-457-826 (1-2109)

QY 246 A1AG1Y1leSerArgerAlathr1le 254
DB 1200 GCGGCATCTCGCGGTCCGCCACCATC 1226

RESULT 21

US-09-614-124B-826
; Sequence 826, Application US/09614124B
; Patent No. 6630574
; GENERAL INFORMATION:
; APPLICANT: Wang, Tonglong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darrick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121.478C9
; CURRENT APPLICATION NUMBER: US/09/614,124B
; CURRENT FILING DATE: 2001-07-11
; NUMBER OF SEQ ID NOS: 1668
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 826
; LENGTH: 2109
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-614-124B-826

Alignment Scores:

Pred. No.:	30.4	Length:	2109
Score:	9.00	Matches:	9
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	1.35%	Indels:	0
DB:	4	Gaps:	0

US-10-029-345A-109 (1-665) x US-09-614-124B-826 (1-2109)

QY 246 A1AG1Y1leSerArgerAlathr1le 254
DB 1200 GCGGCATCTCGCGGTCCGCCACCATC 1226

RESULT 22

US-09-671-325-826
; Sequence 826, Application US/09671325
; Patent No. 6667154
; GENERAL INFORMATION:
; APPLICANT: Wang, Tonglong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.

1200 GCGGCATCTCGCGTCCGCCACCATC 1226

APPLICANT: Ve

APPLICANT: Carter, Darrick

APPLICANT: Retter, Marc
APPLICANT: Mannion, Jane
APPLICANT: Fan, Liqun
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.478C14
CURRENT APPLICATION NUMBER: US/09/702,705
CURRENT FILING DATE: 2000-10-30
NUMBER OF SEQ ID NOS: 1833
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 804
LENGTH: 4637
TYPE: DNA
ORGANISM: Homo sapiens
US-09-702-705-804

Alignment Scores:
Pred. No.: 67.3 Length: 4637
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.35% Indels: 0
DB: 4 Gaps: 0

US-10-029-345A-109 (1-665) x US-09-702-705-804 (1-4637)

QY 246 AAGGYYIleSeraYSerAlaThrIle 254
DB 569 GCGGCATCTCGCGTCGCGCACCATC 595

RESULT 26
US-09-736-457-804
Sequence 804, Application US/09736457
Patent No. 6509448
GENERAL INFORMATION:
APPLICANT: Wang, Tongtong
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Lodes, Michael A.
APPLICANT: Fanger, Gary
APPLICANT: Vedvick, Tom
APPLICANT: Carter, Darick
APPLICANT: Retter, Marc
APPLICANT: Mannion, Jane
APPLICANT: Fan, Liqun
APPLICANT: Wang, Aljun
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.478C15
CURRENT APPLICATION NUMBER: US/09/736,457
CURRENT FILING DATE: 2000-12-13
NUMBER OF SEQ ID NOS: 1864
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 804
LENGTH: 4637
TYPE: DNA
ORGANISM: Homo sapiens
US-09-736-457-804

Alignment Scores:
Pred. No.: 67.3 Length: 4637
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.35% Indels: 0
DB: 4 Gaps: 0

US-10-029-345A-109 (1-665) x US-09-736-457-804 (1-4637)

QY 246 AAGGYYIleSeraYSerAlaThrIle 254
DB 569 GCGGCATCTCGCGTCGCGCACCATC 595

RESULT 27

US-09-614-124B-804
Sequence 804, Application US/09614124B
Patent No. 6630574
GENERAL INFORMATION:
APPLICANT: Wang, Tongtong
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Lodes, Michael A.
APPLICANT: Fanger, Gary
APPLICANT: Vedvick, Tom
APPLICANT: Carter, Darick
APPLICANT: Retter, Marc
APPLICANT: Mannion, Jane
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
FILE REFERENCE: 210121.478C9
CURRENT APPLICATION NUMBER: US/09/614,124B
CURRENT FILING DATE: 2001-07-11
NUMBER OF SEQ ID NOS: 1668
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 804
LENGTH: 4637
TYPE: DNA
ORGANISM: Homo sapiens
US-09-614-124B-804

Alignment Scores:
Pred. No.: 67.3 Length: 4637
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.35% Indels: 0
DB: 4 Gaps: 0

US-10-029-345A-109 (1-665) x US-09-614-124B-804 (1-4637)

QY 246 AAGGYYIleSeraYSerAlaThrIle 254
DB 569 GCGGCATCTCGCGTCGCGCACCATC 595

RESULT 28
US-09-671-325-804
Sequence 804, Application US/09671325
Patent No. 6667154
GENERAL INFORMATION:
APPLICANT: Wang, Tongtong
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Lodes, Michael A.
APPLICANT: Fanger, Gary
APPLICANT: Vedvick, Tom
APPLICANT: Carter, Darick
APPLICANT: Retter, Marc
APPLICANT: Mannion, Jane
APPLICANT: Fan, Liqun
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.478C12
CURRENT APPLICATION NUMBER: US/09/671,325
CURRENT FILING DATE: 2000-09-26
NUMBER OF SEQ ID NOS: 1825
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 804
LENGTH: 4637
TYPE: DNA
ORGANISM: Homo sapiens
US-09-671-325-804

Alignment Scores:
Pred. No.: 67.3 Length: 4637
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.35% Indels: 0
DB: 4 Gaps: 0

US-10-029-345A-109 (1-665) x US-09-671-325-804 (1-4637)
QY 246 AlagylleSerArgSerAlaThrIle 254
DB 569 GCGGCATCTCGCGTCCGCCACCATC 595

RESULT 29
US-09-589-184-804
; Sequence 804, Application US/09589184
; Patent No. 6686447
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Lodes, Michael A.
; APPLICANT: Fanger, Gary
; APPLICANT: Vedvick, Tom
; APPLICANT: Carter, Darick
; APPLICANT: Retter, Marc
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.478C8
; CURRENT APPLICATION NUMBER: US/09/589,184
; NUMBER OF FILING DATE: 2000-06-05
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 804
; LENGTH: 4637
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-589-184-804

Alignment Scores:
Pred. No.: 67.3 length: 4637
Score: 9.00 Matches: 9
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.35% Indels: 0
DB: 4 Gaps: 0

US-10-029-345A-109 (1-665) x US-09-589-184-804 (1-4637)
QY 246 AlagylleSerArgSerAlaThrIle 254
DB 569 GCGGCATCTCGCGTCCGCCACCATC 595

RESULT 30
PCT-US92-08090-3
; Sequence 3, Application PC/TUS9208090
; GENERAL INFORMATION:
; APPLICANT: Tang, J. N.
; TITLE OF INVENTION: Fusion Protein Genes for Treatment of
; TITLE OF INVENTION: Aids
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kilpatrick & Cody
; STREET: 100 Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: U.S.
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/08090
; FILING DATE: 19920922
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.

REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: OMRP129
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-572-6508
TELEFAX: 404-572-6555
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1421 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapien
TISSUE TYPE: Epithelial
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..6
OTHER INFORMATION: /note= "Restriction site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1275..1280
OTHER INFORMATION: /note= "Restriction site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1416..1421
OTHER INFORMATION: /note= "Restriction site"
PCT-US92-08090-3

Alignment Scores:
Pred. No.: 219 length: 1421
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 5 Gaps: 0

US-10-029-345A-109 (1-665) x PCT-US92-08090-3 (1-1421)
QY 605 AsparaGAlaAspSerArgSer 612
DB 312 GATCGCGCTGACTCAGAGAGAGC 335

RESULT 31
US-08-284-465-2
; Sequence 2, Application US/08284465
; Patent No. 5830457
; GENERAL INFORMATION:
; APPLICANT: GICQUEL, BRIGITTE
; APPLICANT: TIMM, JULIANO
; APPLICANT: TRIAS, JOAQUIM
; APPLICANT: DUEZ, COLETTE
; APPLICANT: PERILLI, MARIA-GRAZIA
; APPLICANT: DUSART, JEAN
; APPLICANT: FREBE, JEAN-MARIE
; TITLE OF INVENTION: RECOMBINANT BETA-LACTAMASE, USABLE AS
; TITLE OF INVENTION: CARRIER MOLECULE FOR THE PREPARATION OF IMMUNOGENIC
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/284,465
FILING DATE: 14-NOV-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA: FR 9201713
APPLICATION NUMBER: FR 9201713
FILING DATE: 14-FEB-1992
PRIOR APPLICATION DATA: PCT/FR 93/00151
APPLICATION NUMBER: PCT/FR 93/00151
FILING DATE: 12-FEB-1993
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1426 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 395..1276
US-08-284-465-2

Alignment Scores:
Pred. No.: 220 Length: 1426
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 2 Gaps: 0

US-10-029-345a-109 (1-665) x US-08-284-465-2 (1-1426)

OY 549 SerTh-ProserLeuthrSerSer 556
DB 268 TCNACGCCGCTCCTCACCACTTCG 291

RESULT 32
US-08-284-465-7
Sequence 7, Application US/08284465
Patent No. 5830457
GENERAL INFORMATION:
APPLICANT: GICQUEL, BRIGITTE
APPLICANT: TIMM, JULIANO
APPLICANT: TRIAS, JOAQUIM
APPLICANT: DUEZ, COLETTE
APPLICANT: PERILLI, MARIA-GRAZIA
APPLICANT: DUSART, JEAN
APPLICANT: FRERE, JEAN-MARIE
TITLE OF INVENTION: RECOMBINANT BETA-LACTAMASE, USABLE AS
TITLE OF INVENTION: CARRIER MOLECULE FOR THE PREPARATION OF IMMUNOGENIC
TITLE OF INVENTION: COMPOSITIONS
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/284,465
FILING DATE: 14-NOV-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:

APPLICATION NUMBER: FR 9201713
FILING DATE: 14-FEB-1992
PRIOR APPLICATION DATA: PCT/FR 93/00151
APPLICATION NUMBER: PCT/FR 93/00151
FILING DATE: 12-FEB-1993
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1426 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 395..1276
FEATURE:
NAME/KEY: misc_feature
LOCATION: 563..568
FEATURE:
NAME/KEY: misc_feature
LOCATION: 672..677
FEATURE:
NAME/KEY: misc_feature
LOCATION: 506..532
FEATURE:
NAME/KEY: misc_feature
LOCATION: 539..562
FEATURE:
NAME/KEY: misc_feature
LOCATION: 572..589
FEATURE:
NAME/KEY: misc_feature
LOCATION: 620..652
FEATURE:
NAME/KEY: misc_feature
LOCATION: 725..748
FEATURE:
NAME/KEY: misc_feature
LOCATION: 764..790
FEATURE:
NAME/KEY: misc_feature
LOCATION: 815..835
FEATURE:
NAME/KEY: misc_feature
LOCATION: 842..874
FEATURE:
NAME/KEY: misc_feature
LOCATION: 905..925
FEATURE:
NAME/KEY: misc_feature
LOCATION: 956..988
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1016..1036
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1052..1075
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1061..1066
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1091..1117
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1127..1156
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1172..1204

```
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1235..1276
US-08-284-465-7

Alignment Scores:
Pred. No.: 220      Length: 1426
Score: 8.00      Matches: 8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 1.20%      Indels: 0
DB: 2      Gaps: 0

US-10-029-345A-109 (1-665) x US-08-284-465-7 (1-1426)
QY 549 SerThpProSerLeuThrSerSer 556
DB 268 TCACGCCGCTGCTCACAGTTCC 291

RESULT 33
PCT-US92-08090-2
Sequence 2, Application PC/TUS9208090
GENERAL INFORMATION:
APPLICANT: Tang, J. N.
TITLE OF INVENTION: Fusion Protein Genes for Treatment of
TITLE OF INVENTION: Aids
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kilpatrick & Cody
STREET: 100 Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: U.S.
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/08090
FILING DATE: 19920922
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: OMRP129
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-572-6508
TELEFAX: 404-572-6555
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1448 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Homo sapien
TISSUE TYPE: Epithelial
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..6
OTHER INFORMATION: /note= "Restriction site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1275..1280
OTHER INFORMATION: /note= "Restriction site"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1444..1448
```

```
OTHER INFORMATION: /note= "Restriction site"
PCT-US92-08090-2

Alignment Scores:
Pred. No.: 224      Length: 1448
Score: 8.00      Matches: 8
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match: 1.20%      Indels: 0
DB: 5      Gaps: 0

US-10-029-345A-109 (1-665) x PCT-US92-08090-2 (1-1448)
QY 605 AspaGAlaAapSerArgArgSer 612
DB 312 GATCGCGCTGACTCAGAGAAAGC 335

RESULT 34
US-08-236-311-5
Sequence 5, Application US/08236311
Patent No. 556335
GENERAL INFORMATION:
APPLICANT: Capon, Daniel J.
APPLICANT: Gregory, Timothy J.
TITLE OF INVENTION: Adhesion Variants
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/236,311
FILING DATE: 02-MAY-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/936190
FILING DATE: 26-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/842777
FILING DATE: 18-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/250785
FILING DATE: 28-SEP-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/104329
FILING DATE: 02-OCT-1987
ATTORNEY/AGENT INFORMATION:
NAME: Haasek, Janet E.
REGISTRATION NUMBER: 28,616
REFERENCE/DOCKET NUMBER: 444P1C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1896
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1508 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-236-311-5

Alignment Scores:
Pred. No.: 233      Length: 1508
Score: 8.00      Matches: 8
```

Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
Gaps: 0
DB: 1
US-10-029-345a-109 (1-665) x US-08-236-311-5 (1-1508)
QY 605 AsparaGalaSpSerArgSer 612
DB 403 GATCGCGTGACTCAAGAAGAGC 426
RESULT 35
US-08-457-918-5
Sequence 5, Application US/08457918
Patent No. 6117655
GENERAL INFORMATION:
APPLICANT: Capon, Daniel J.
APPLICANT: Gregory, Timothy J.
TITLE OF INVENTION: Adhesion Variants
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/457,918
FILING DATE: 1-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/236311
FILING DATE: 02-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/936190
FILING DATE: 26-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/842777
FILING DATE: 18-FEB-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/250785
FILING DATE: 28-SEP-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/104329
FILING DATE: 02-OCT-1987
ATTORNEY/AGENT INFORMATION:
NAME: Kubiñec, Jeffrey S.
REGISTRATION NUMBER: 36,575
REFERENCE/DOCKET NUMBER: P0444P1C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-8228
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1508 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-457-918-5
Alignment Scores:
Pred. No.: 233 Length: 1508
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0

DB: 3 Gaps: 0
US-10-029-345a-109 (1-665) x US-08-457-918-5 (1-1508)
QY 605 AsparaGalaSpSerArgSer 612
DB 403 GATCGCGTGACTCAAGAAGAGC 426
RESULT 36
US-08-417-495-3
Sequence 3, Application US/08417495
Patent No. 5843728
GENERAL INFORMATION:
APPLICANT: Seed, Brian et al.
TITLE OF INVENTION: Redirection of Cellular Immunity by
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: Wordperfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/417,495
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/203,866
FILING DATE:
APPLICATION NUMBER: US/07/847,566
FILING DATE:
APPLICATION NUMBER: 07/665,961
FILING DATE: March 7, 1991
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/119002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1599 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-417-495-3
Alignment Scores:
Pred. No.: 247 Length: 1599
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
Gaps: 0
DB: 2
US-10-029-345a-109 (1-665) x US-08-417-495-3 (1-1599)
QY 605 AsparaGalaSpSerArgSer 612
DB 232 GATCGCGTGACTCAAGAAGAGC 255
RESULT 37

US-08-284-391B-3
; Sequence 3, Application US/08284391B
; Patent No. 5851828
; GENERAL INFORMATION:
; APPLICANT: Seed, Brian
; APPLICANT: Banapour, Babak
; APPLICANT: Romeo, Charles
; APPLICANT: Kolanus, Waldemar
; TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
; TITLE OF INVENTION: CELLS BY CHIMERIC CD4 RECEPTOR-BEARING CELLS
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Elbing LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/284,391B
; FILING DATE: 02-AUG-1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/195,395
; FILING DATE: 14-FEB-1994
; APPLICATION NUMBER: 07/847,566
; FILING DATE: 06-MAR-1992
; APPLICATION NUMBER: 07/665,961
; FILING DATE: 07-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Elbing, Karen L.
; REGISTRATION NUMBER: 35,238
; REFERENCE/DOCKET NUMBER: 00786/247001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-428-0200
; TELEFAX: 617-428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1599 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-284-391B-3
Alignment Scores:
Pred. No.: 247
Score: 8.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 1.20%
Matches: 1599
Conservative: 8
Mismatch: 0
Indels: 0
Gaps: 0
DB: 2
US-10-029-345A-109 (1-665) x US-08-284-391B-3 (1-1599)
QY 605 AspaGAlaSpSerArgSer 612
DB 232 GATCGCGCTGACTCAAGAAAGAGC 255
RESULT 38
US-09-218-950-3
; Sequence 3, Application US/09218950
; Patent No. 6284240
; GENERAL INFORMATION:
; APPLICANT: Seed, Brian
; APPLICANT: Banapour, Babak
; APPLICANT: Romeo, Charles

APPLICANT: Kolanus, Waldemar
; TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED
; TITLE OF INVENTION: CELLS BY CHIMERIC CD4 RECEPTOR-BEARING CELLS
; NUMBER OF SEQUENCES: 53
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Elbing LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/218,950
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/284,391
; FILING DATE: 02-AUG-1994
; APPLICATION NUMBER: 08/195,395
; FILING DATE: 14-FEB-1994
; APPLICATION NUMBER: 07/847,566
; FILING DATE: 06-MAR-1992
; APPLICATION NUMBER: 07/665,961
; FILING DATE: 07-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Elbing, Karen L.
; REGISTRATION NUMBER: 35,238
; REFERENCE/DOCKET NUMBER: 00786/247001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-428-0200
; TELEFAX: 617-428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1599 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-218-950-3
Alignment Scores:
Pred. No.: 247
Score: 8.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 1.20%
Matches: 1599
Conservative: 8
Mismatch: 0
Indels: 0
Gaps: 0
DB: 3
US-10-029-345A-109 (1-665) x US-09-218-950-3 (1-1599)
QY 605 AspaGAlaSpSerArgSer 612
DB 232 GATCGCGCTGACTCAAGAAAGAGC 255
RESULT 39
PCT-US92-01785-3
; Sequence 3, Application PC/TUS9201785
; GENERAL INFORMATION:
; APPLICANT: The General Hospital Corporation
; TITLE OF INVENTION: Redirection of Cellular Immunity by Receptor
; TITLE OF INVENTION: Chimeras
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA

COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: Wordperfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/01785
FILING DATE: 19920306
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/665,961
FILING DATE: March 7, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/119002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1599 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PCT-US92-01785-3

Alignment Scores:
Pred. No.: 247 Length: 1599
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 5 Gaps: 0

US-10-029-345A-109 (1-665) x PCT-US92-01785-3 (1-1599)

QY 605 AsparGAlaasparGArgSer 612
DB 232 GATCGCGCTGACTCAAGAGAGC 255

RESULT 40
PCT-US95-00454-3
Sequence 3, Application PC/TUS9500454
GENERAL INFORMATION:
APPLICANT: Seed, Brian et al.
TITLE OF INVENTION: Targeted Cytolysis of HIV-Infected
TITLE OF INVENTION: Cells by Chimeric CD4 Receptor-
TITLE OF INVENTION: Bearing Cells
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: Wordperfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/00454
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/847,566
FILING DATE: March 6, 1992

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/665,961
FILING DATE: March 7, 1991
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/247001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1599 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PCT-US93-00454-3

Alignment Scores:
Pred. No.: 247 Length: 1599
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 5 Gaps: 0

US-10-029-345A-109 (1-665) x PCT-US95-00454-3 (1-1599)

QY 605 AsparGAlaasparGArgSer 612
DB 232 GATCGCGCTGACTCAAGAGAGC 255

RESULT 41
US-08-916-902A-2/C
Sequence 2, Application US/08916902A
Patent No. 5871930
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HIGH AFFINITY IMMUNOGLOBULIN E
TITLE OF INVENTION: RECEPTOR-LIKE PROTEIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: PASTESQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/916,902A
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0371 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166

TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1669 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRAINOT04
; CLONE: 927955
US-08-916-902A-2

Alignment Scores:
Pred. No.: 258 Length: 1669
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 2 Gaps: 0

US-10-029-345A-109 (1-665) x US-08-916-902A-2 (1-1669)

Qy 137 GIUGLYSserThreValPro 144
Db 1349 GAAGGAAAAAGCATTGGTCCCC 1326

RESULT 42
US-09-213-389-2/c
; Sequence 2, Application US/09213389
; Patent No. 5977072
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: HIGH AFFINITY IMMUNOGLOBULIN E
; TITLE OF INVENTION: RECEPTOR-LIKE PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/213,389
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/916,902
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0371 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1669 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRAINOT04
; CLONE: 927955

US-09-213-389-2

Alignment Scores:
Pred. No.: 258 Length: 1669
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 2 Gaps: 0

US-10-029-345A-109 (1-665) x US-09-213-389-2 (1-1669)

Qy 137 GIUGLYSserThreValPro 144
Db 1349 GAAGGAAAAAGCATTGGTCCCC 1326

RESULT 43
US-08-417-495-1
; Sequence 1, Application US/08417495
; Patent No. 5843728
; GENERAL INFORMATION:
; APPLICANT: Seed, Brian et al.
; TITLE OF INVENTION: Redirection of Cellular Immunity by
; TITLE OF INVENTION: Receptor
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; COMPUTER: IBM PS/2 Model 502 or 55SX
; OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
; SOFTWARE: Wordperfect (Version 5.0)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/417,495
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/203,866
; FILING DATE:
; APPLICATION NUMBER: US/07/847,566
; FILING DATE:
; APPLICATION NUMBER: 07/665,961
; FILING DATE: March 7, 1991
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Paul T.
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 00786/119002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1728 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-417-495-1

Alignment Scores:
Pred. No.: 267 Length: 1728
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 2 Gaps: 0

Chimeras

US-10-029-345A-109 (1-665) x US-08-417-495-1 (1-1728)

QY 605 AaPaRgAlaAsPSeRrGArGSeR 612

DB 232 GATCGCGCTGACTCAAGAAGAGC 255

RESULT 44
US-08-284-391B-1

/ Sequence 1, Application US/08284391B

/ Patent No. 5851828

/ GENERAL INFORMATION:

/ APPLICANT: Seed, Brian

/ APPLICANT: Banapour, Babak

/ APPLICANT: Romeo, Charles

/ APPLICANT: Kolanus, Waldemar

/ TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED

/ CELLS BY CHIMERIC CD4 RECEPTOR- BEARING CELLS

/ NUMBER OF SEQUENCES: 53

/ CORRESPONDENCE ADDRESS:

/ ADDRESSEE: Clark & Elbing LLP

/ STREET: 176 Federal Street

/ CITY: Boston

/ STATE: MA

/ COUNTRY: USA

/ ZIP: 02110

/ COMPUTER READABLE FORM:

/ MEDIUM TYPE: Diskette

/ COMPUTER: IBM Compatible

/ OPERATING SYSTEM: DOS

/ SOFTWARE: FASTSEQ for Windows Version 2.0

/ CURRENT APPLICATION DATA:

/ APPLICATION NUMBER: US/08/284,391B

/ FILING DATE: 02-AUG-1994

/ CLASSIFICATION: 514

/ PRIOR APPLICATION DATA:

/ APPLICATION NUMBER: 08/195,395

/ FILING DATE: 14-FEB-1994

/ APPLICATION NUMBER: 07/847,566

/ FILING DATE: 06-MAR-1992

/ APPLICATION NUMBER: 07/665,961

/ FILING DATE: 07-MAR-1991

/ ATTORNEY/AGENT INFORMATION:

/ NAME: Elbing, Karen L.

/ REGISTRATION NUMBER: 35,238

/ REFERENCE/DOCKET NUMBER: 00786/247001

/ TELECOMMUNICATION INFORMATION:

/ TELEPHONE: 617-428-0200

/ TELEFAX: 617-428-7045

/ TELEX:

/ INFORMATION FOR SEQ ID NO: 1:

/ SEQUENCE CHARACTERISTICS:

/ LENGTH: 1728 base pairs

/ TYPE: nucleic acid

/ STRANDEDNESS: double

/ TOPOLOGY: linear

/ MOLECULE TYPE: cDNA

/ US-08-284-391B-1

/ Alignment Scores:

/ Pred. No.: 267

/ Score: 8.00

/ Percent Similarity: 100.00%

/ Best Local Similarity: 100.00%

/ Query Match: 1.20%

/ DB: 2

/ Gaps: 0

/ Length: 1728

/ Matches: 8

/ Conservative: 0

/ Mismatches: 0

/ Indels: 0

/ Gaps: 0

/ US-10-029-345A-109 (1-665) x US-08-284-391B-1 (1-1728)

RESULT 45
US-09-218-950-1

/ Sequence 1, Application US/09218950

/ Patent No. 6284240

/ GENERAL INFORMATION:

/ APPLICANT: Seed, Brian

/ APPLICANT: Banapour, Babak

/ APPLICANT: Romeo, Charles

/ APPLICANT: Kolanus, Waldemar

/ TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED

/ CELLS BY CHIMERIC CD4 RECEPTOR- BEARING CELLS

/ NUMBER OF SEQUENCES: 53

/ CORRESPONDENCE ADDRESS:

/ ADDRESSEE: Clark & Elbing LLP

/ STREET: 176 Federal Street

/ CITY: Boston

/ STATE: MA

/ COUNTRY: USA

/ ZIP: 02110

/ COMPUTER READABLE FORM:

/ MEDIUM TYPE: Diskette

/ COMPUTER: IBM Compatible

/ OPERATING SYSTEM: DOS

/ SOFTWARE: FASTSEQ for Windows Version 2.0

/ CURRENT APPLICATION DATA:

/ APPLICATION NUMBER: US/09/218,950

/ FILING DATE:

/ CLASSIFICATION:

/ PRIOR APPLICATION DATA:

/ APPLICATION NUMBER: US/08/284,391

/ FILING DATE: 02-AUG-1994

/ APPLICATION NUMBER: 08/195,395

/ FILING DATE: 14-FEB-1994

/ APPLICATION NUMBER: 07/847,566

/ FILING DATE: 06-MAR-1992

/ APPLICATION NUMBER: 07/665,961

/ FILING DATE: 07-MAR-1991

/ ATTORNEY/AGENT INFORMATION:

/ NAME: Elbing, Karen L.

/ REGISTRATION NUMBER: 35,238

/ REFERENCE/DOCKET NUMBER: 00786/247001

/ TELECOMMUNICATION INFORMATION:

/ TELEPHONE: 617-428-0200

/ TELEFAX: 617-428-7045

/ TELEX:

/ INFORMATION FOR SEQ ID NO: 1:

/ SEQUENCE CHARACTERISTICS:

/ LENGTH: 1728 base pairs

/ TYPE: nucleic acid

/ STRANDEDNESS: double

/ TOPOLOGY: linear

/ MOLECULE TYPE: cDNA

/ US-09-218-950-1

/ Alignment Scores:

/ Pred. No.: 267

/ Score: 8.00

/ Percent Similarity: 100.00%

/ Best Local Similarity: 100.00%

/ Query Match: 1.20%

/ DB: 3

/ Gaps: 0

/ Length: 1728

/ Matches: 8

/ Conservative: 0

/ Mismatches: 0

/ Indels: 0

/ Gaps: 0

/ US-10-029-345A-109 (1-665) x US-09-218-950-1 (1-1728)

RESULT 46
PCT-US92-01785-1

/ Sequence 1, Application PC/TUS9201785

/ GENERAL INFORMATION:

/ APPLICANT: The General Hospital Corporation

/ APPLICANT: Banapour, Babak

/ APPLICANT: Romeo, Charles

/ APPLICANT: Kolanus, Waldemar

/ TITLE OF INVENTION: TARGETED CYTOLYSIS OF HIV-INFECTED

/ CELLS BY CHIMERIC CD4 RECEPTOR- BEARING CELLS

/ NUMBER OF SEQUENCES: 53

/ CORRESPONDENCE ADDRESS:

/ ADDRESSEE: Clark & Elbing LLP

/ STREET: 176 Federal Street

/ CITY: Boston

/ STATE: MA

/ COUNTRY: USA

/ ZIP: 02110

/ COMPUTER READABLE FORM:

/ MEDIUM TYPE: Diskette

/ COMPUTER: IBM Compatible

/ OPERATING SYSTEM: DOS

/ SOFTWARE: FASTSEQ for Windows Version 2.0

/ CURRENT APPLICATION DATA:

/ APPLICATION NUMBER: US/09/218,950

/ FILING DATE:

/ CLASSIFICATION:

/ PRIOR APPLICATION DATA:

/ APPLICATION NUMBER: US/08/284,391

/ FILING DATE: 02-AUG-1994

/ APPLICATION NUMBER: 08/195,395

/ FILING DATE: 14-FEB-1994

/ APPLICATION NUMBER: 07/847,566

/ FILING DATE: 06-MAR-1992

/ APPLICATION NUMBER: 07/665,961

/ FILING DATE: 07-MAR-1991

/ ATTORNEY/AGENT INFORMATION:

/ NAME: Elbing, Karen L.

/ REGISTRATION NUMBER: 35,238

/ REFERENCE/DOCKET NUMBER: 00786/247001

/ TELECOMMUNICATION INFORMATION:

/ TELEPHONE: 617-428-0200

/ TELEFAX: 617-428-7045

/ TELEX:

/ INFORMATION FOR SEQ ID NO: 1:

/ SEQUENCE CHARACTERISTICS:

/ LENGTH: 1728 base pairs

/ TYPE: nucleic acid

/ STRANDEDNESS: double

/ TOPOLOGY: linear

/ MOLECULE TYPE: cDNA

/ US-09-218-950-1

/ Alignment Scores:

/ Pred. No.: 267

/ Score: 8.00

/ Percent Similarity: 100.00%

/ Best Local Similarity: 100.00%

/ Query Match: 1.20%

/ DB: 3

/ Gaps: 0

/ Length: 1728

/ Matches: 8

/ Conservative: 0

/ Mismatches: 0

/ Indels: 0

/ Gaps: 0

/ US-10-029-345A-109 (1-665) x US-09-218-950-1 (1-1728)

TITLE OF INVENTION: Redirection of Cellular Immunity by Receptor
TITLE OF INVENTION: Chimeras
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)
SOFTWARE: Wordperfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/01785
FILING DATE: 19920306
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/665,961
FILING DATE: March 7, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/119002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1728 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PCT-US92-01785-1

Alignment Scores:
Pred. No.: 267 Length: 1728
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 5 Gaps: 0

US-10-029-345A-109 (1-665) x PCT-US92-01785-1 (1-1728)

QY 605 AspaGAlaaspSerArgSer 612
Db 232 GATCGCGCTGACTCAAGAAAGAC 255

RESULT 47
PCT-US95-00454-1
Sequence 1, Application PC/TUS9500454
GENERAL INFORMATION:
APPLICANT: Seed, Brian et al.
TITLE OF INVENTION: Targeted Cytolysis of HIV-Infected
TITLE OF INVENTION: Cells by Chimeric CD4 Receptor-
TITLE OF INVENTION: Bearing Cells
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
COMPUTER: IBM PS/2 Model 502 or 55SX
OPERATING SYSTEM: IBM P.C. DOS (Version 3.30)

SOFTWARE: Wordperfect (Version 5.0)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/00454
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/847,566
FILING DATE: March 6, 1992
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/665,961
FILING DATE: March 7, 1991
ATTORNEY/AGENT INFORMATION:
NAME: Clark, Paul T.
REGISTRATION NUMBER: 30,162
REFERENCE/DOCKET NUMBER: 00786/247001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-5070
TELEFAX: (617) 542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1728 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PCT-US95-00454-1

Alignment Scores:
Pred. No.: 267 Length: 1728
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 5 Gaps: 0

US-10-029-345A-109 (1-665) x PCT-US95-00454-1 (1-1728)

QY 605 AspaGAlaaspSerArgSer 612
Db 232 GATCGCGCTGACTCAAGAAAGAC 255

RESULT 48
US-08-466-368-3
Sequence 3, Application US/08466368
GENERAL INFORMATION:
APPLICANT: Maddan, Paul J.
APPLICANT: Littman, Dan R.
APPLICANT: Chess, Leonard
APPLICANT: Axel, Richard
APPLICANT: Weiss, Robin
APPLICANT: McDougal, J. S.
TITLE OF INVENTION: DNA ENCODING THE T CELL SURFACE PROTEIN
TITLE OF INVENTION: T4 AND USE OF FRAGMENTS OF T4 IN THE TREATMENT OF AIDS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,368
FILING DATE: 06-JUN-1995

```
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 24577-F1-B/JPM/AKC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-278-0400
; TELEFAX: 212-391-0525
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1742 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 76..1449
;
US-08-466-368-3
;
Alignment Scores:
Pred. No.: 269 Length: 1742
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 3 Gaps: 0

US-10-029-345A-109 (1-665) x US-08-466-368-3 (1-1742)
QY 605 AsparGAlaaspSerArgSer 612
DB 307 GATCGCGCTGACTCAAGAGAAGC 330

RESULT 49
US-09-517-605-7
; Sequence 7, Application US/09517605
; Patent No. 6391567
; GENERAL INFORMATION:
; APPLICANT: Littman, Dan R.
; APPLICANT: Kwon, Douglas S.
; APPLICANT: van Kooyk, Yvette
; APPLICANT: Geijtenbeck, Theo
; TITLE OF INVENTION: METHODS OF USING A FACILITATOR OF RETROVIRAL ENTRY INTO
; FILE REFERENCE: 1049-1-017
; CURRENT APPLICATION NUMBER: US/09/517,605
; CURRENT FILING DATE: 2000-03-02
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 1742
; TYPE: DNA
; ORGANISM: Homo sapiens
;
US-09-517-605-7
;
Alignment Scores:
Pred. No.: 269 Length: 1742
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 4 Gaps: 0

US-10-029-345A-109 (1-665) x US-09-517-605-7 (1-1742)
QY 605 AsparGAlaaspSerArgSer 612
DB 307 GATCGCGCTGACTCAAGAGAAGC 330

RESULT 50
US-08-470-998-1
; Sequence 1, Application US/08470998
```

```
; Patent No. 6570000
; GENERAL INFORMATION:
; APPLICANT: Maddon, Paul J.
; APPLICANT: Littman, Dan R.
; APPLICANT: Chess, Leonard
; APPLICANT: Axel, Richard
; APPLICANT: Weiss, Robin
; APPLICANT: McDougal, J. S.
; TITLE OF INVENTION: DNA ENCODING THE T CELL SURFACE PROTEIN
; TITLE OF INVENTION: T4 AND USE OF FRAGMENTS OF T4 IN THE TREATMENT OF AIDS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Cooper & Dunham LLP
; STREET: 1185 Avenue of Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
;
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,998
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 24577-F1-B/JPM/AKC
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
;
INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1742 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: YES
; ANTI-SENSE: YES
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 76..1449
;
US-08-470-998-1
;
Alignment Scores:
Pred. No.: 269 Length: 1742
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.20% Indels: 0
DB: 4 Gaps: 0

US-10-029-345A-109 (1-665) x US-08-470-998-1 (1-1742)
QY 605 AsparGAlaaspSerArgSer 612
DB 307 GATCGCGCTGACTCAAGAGAAGC 330

Search completed: June 22, 2004, 10:11:37
Job time : 156 secs
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 14:11:03 ; Search time 60 Seconds
(without alignments)
3131.566 Million cell updates/sec

Title: US-10-029-345A-109

Sequence: 1 MAHEMIGTQIVTERLVALL...LGVKSQSGSFGSMETIEVS 665

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1586107 seqs, 282547505 residues

Word size : 473

Total number of hits satisfying chosen parameters: 2

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 75 summaries

Database :

- 1: A_Geneseq_299a04:*
- 2: geneseqp1980s:*
- 3: geneseqp1990s:*
- 4: geneseqp2000s:*
- 5: geneseqp2001s:*
- 6: geneseqp2002s:*
- 7: geneseqp2003as:*
- 8: geneseqp2003bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	665	100.0	665	5 ABR52381	Ab152381 Protein r
2	665	100.0	665	5 ABR52407	Ab152407 Protein r

ALIGNMENTS

RESULT 1
ID ABR52381 standard; protein; 665 AA.
XX
AC ABR52381;
XX
DT 19-JUN-2003 (first entry)
DE Protein relating to the invention SEQ ID NO: 109.
XX
XX anti-proliferative; hepatotropic; nephrotropic; anti-arthritic;
KM anti-proliferative; cardiant; cytosolic; gene therapy; liver disease;
KM proliferative disorder; renal failure; cardiovascular disorder;
KM immunological disorder; arthritis; psoriasis; congenital heart defect;
KM congestive heart failure; cancer; human; mouse; yeast; rat; fruitfly.
XX
OS Homo sapiens.
XX

PN	MO200257460-A2.	
XX		
PD	25-JUL-2002.	
XX		
PF	20-DEC-2001; 2001WO-US050459.	
XX		
PR	20-DEC-2000; 2000US-0256868P.	
PR	30-MAR-2001; 2001US-0280186P.	
PR	01-MAY-2001; 2001US-0287735P.	
PR	05-JUN-2001; 2001US-0295848P.	
PR	25-JUN-2001; 2001US-0300465P.	
PA	(BRIM) BRISTOL-MYERS SQUIBB CO.	
XX		
PI	Jackson DG, Feder J, Nelson T, Mintier G, Ramanathan C, Lee L;	
PI	Stemers N, Bol D, Schieven G, Finger J, Todderud CG, Basolino D;	
PI	Kryslek S, Mcatee P, Suchard S, Banas D;	
DR	WPI; 2002-599721/64.	
DR	N-PSDB; ACC60559.	
XX		
PT	Novel polynucleotides encoding human phosphatase polypeptides useful in	
PT	the prevention or treatment of e.g. proliferative and cardiovascular	
PT	disorders.	
XX		
PS	Claim 5; Fig 12; 801pp; English.	
XX		
CC	The invention relates to a novel isolated nucleic acid comprising a	
CC	polynucleotide having a nucleotide sequence selected from 40	
CC	polynucleotides fully defined in the specification. The polynucleotide of	
CC	the invention has anti-proliferative, hepatotropic, nephrotropic,	
CC	anti-arthritic, anti-proliferative, cardiant, and cytosolic activity. The	
CC	polynucleotide may have a use in gene therapy. A polynucleotide or	
CC	polypeptide of the invention is useful for preventing, treating or	
CC	ameliorating a medical condition, e.g. a proliferative disorder. They are	
CC	also useful for treating e.g. liver disease, renal failure, immunological	
CC	disorders including arthritis and psoriasis, cardiovascular disorders	
CC	such as congenital heart defects and congestive heart failure, and	
CC	cancer. A method of the invention is useful for diagnosing a pathological	
CC	condition or susceptibility to a condition in a subject. The present	
CC	sequence is used in the exemplification of the invention	
XX		
SO	Sequence 665 AA;	
	Query Match 100.0%; Score 665; DB 5; Length 665;	
	Best Local Similarity 100.0%; Pred. No. 0;	
	Matches 665; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 MAHEMIGTQIVTERLVALL...LGVKSQSGSFGSMETIEVS 60	
DB	1 MAHEMIGTQIVTERLVALL...LGVKSQSGSFGSMETIEVS 60	
QY	61 DKVLTTELIGHAKHKVDIDCSQKVVVYDQSSQDVASLSDCF... 120	
DB	61 DKVLTTELIGHAKHKVDIDCSQKVVVYDQSSQDVASLSDCF... 120	
QY	121 LAGGFAEFRCPPGLCEGKSTLVPTCISQPCLPVANI... 180	
DB	121 LAGGFAEFRCPPGLCEGKSTLVPTCISQPCLPVANI... 180	
QY	181 QONGIGYVUNASVTCPPKPFIPESHFLRPVNDSP... 240	
DB	181 QONGIGYVUNASVTCPPKPFIPESHFLRPVNDSP... 240	
QY	241 LVHCLAGISRSATIAIAYMKRMDSLDAVYFV... 300	
DB	241 LVHCLAGISRSATIAIAYMKRMDSLDAVYFV... 300	
QY	301 QTGASGPKSKLKLHLKEXPNEVPVAVSEGGQSE... 360	
DB	301 QTGASGPKSKLKLHLKEXPNEVPVAVSEGGQSE... 360	
QY	361 PSVPSVQPSLLEDSPLVQALSGHLHSADRL... 420	

```
DB 361 PSVPSVQPSLLEDSPLVQALSGHLISADRLBDSNKLKRSFSLDIKSVSYSASMAASLHGF 420
QY 421 SSSDALLEYKPSSTLLDGTNKLCOFSPVOELSEQTPETSPDKKASIPKKLQTARPSDSQ 480
DB 421 SSSDALLEYKPSSTLLDGTNKLCOFSPVOELSEQTPETSPDKKASIPKKLQTARPSDSQ 480
QY 481 SKRLHSVTRSSSGTQORSLISPLHRSQSVEDNYHTSPLFGLSTSQOHLTKSAGLKGWH 540
DB 481 SKRLHSVTRSSSGTQORSLISPLHRSQSVEDNYHTSPLFGLSTSQOHLTKSAGLKGWH 540
QY 541 SDILAPQSTPSTLSSWYFATESSHFYASAIYGSASYSAYSCSGLPTCGDQVYSVRRR 600
DB 541 SDILAPQSTPSTLSSWYFATESSHFYASAIYGSASYSAYSCSGLPTCGDQVYSVRRR 600
QY 601 QKPSDRAISRHWHSPEKQPKRRSCQMEFGESIMSENRREELGKVGSSSFGSME 660
DB 601 QKPSDRAISRHWHSPEKQPKRRSCQMEFGESIMSENRREELGKVGSSSFGSME 660
QY 661 IIEVS 665
DB 661 IIEVS 665

RESULT 2
ABRS2407
ID ABR52407 standard; protein; 665 AA.
XX
AC ABR52407;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein relating to the invention SEQ ID NO: 148.
XX
KW antiproliferative; hepatotropic; nephrotropic; antiarthritic;
KW antipsoriatic; cardiant; cytostatic; gene therapy; liver disease;
KW proliferative disorder; renal failure; cardiovascular disorder;
KW immunological disorder; arthritis; psoriasis; congenital heart defect;
KW congestive heart failure; cancer; human; mouse; yeast; rat; fruitfly.
XX
OS Homo sapiens.
XX
PN MO200257460-A2.
XX
PD 25-JUL-2002.
XX
PF 20-DEC-2001; 2001WO-US050459.
XX
PR 20-DEC-2000; 2000US-0256868P.
PR 30-MAR-2001; 2001US-0280186P.
PR 01-MAY-2001; 2001US-0287735P.
PR 05-JUN-2001; 2001US-0295848P.
PR 25-JUN-2001; 2001US-0300465P.
XX
PA (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
PI Jackson DG, Feder J, Nelson T, Mintier G, Ramanathan C, Lee L;
PI Stiemers N, Bol D, Schieven G, Finger J, Todderud CG, Bassolino D;
PI Krysstek S, Mcatee P, Suchard S, Banas D;
XX
XX WPI; 2002-599721/64.
DR N-PSDB; ACC60572.
PT Novel polynucleotides encoding human phosphatase polypeptides useful in
PT the prevention or treatment of e.g. proliferative and cardiovascular
PT disorders.
XX
XX Disclosure; Fig 19; 801pp; English.
XX
XX The invention relates to a novel isolated nucleic acid comprising a
CC polynucleotide having a nucleotide sequence selected from 40
CC polynucleotides fully defined in the specification. The polynucleotide of
CC the invention has antiproliferative, hepatotropic, nephrotropic,
```

```
CC antiarthritic, antipsoriatic, cardiant, and cytostatic activity. The
CC polynucleotide may have a use in gene therapy. A polynucleotide or
CC polypeptide of the invention is useful for preventing, treating or
CC ameliorating a medical condition, e.g. a proliferative disorder. They are
CC also useful for treating e.g. liver disease, renal failure, immunological
CC disorders including arthritis and psoriasis, cardiovascular disorders
CC such as congenital heart defects and congestive heart failure, and
CC cancer. A method of the invention is useful for diagnosing a pathological
CC condition or susceptibility to a condition in a subject. The present
CC sequence is used in the exemplification of the invention
XX
SQ Sequence 665 AA;
Query Match 100.0%; Score 665; DB 5; Length 665;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 665; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAHEMIGQIVTERIVALLSGTEKVLIDSRPFVEYNTSHLEAININCSKLMRRRLQ 60
DB 1 MAHEMIGQIVTERIVALLSGTEKVLIDSRPFVEYNTSHLEAININCSKLMRRRLQ 60
QY 61 DKVLITTELIQHSAAKHVDIDCSQKVYVYDQSSQDVASISSDCEFTVLLGKLEKSTNSVHL 120
DB 61 DKVLITTELIQHSAAKHVDIDCSQKVYVYDQSSQDVASISSDCEFTVLLGKLEKSTNSVHL 120
QY 121 IAGGFAEFSRCPFGCEGKSTIVPTCISQPCLVANIGPTRILPVLXGCGQDVYANKELI 180
DB 121 IAGGFAEFSRCPFGCEGKSTIVPTCISQPCLVANIGPTRILPVLXGCGQDVYANKELI 180
QY 181 QONGIGYVLNMSYTCPKPDFIPESHFLRPVNDSECEKILPWLIDKSVDFIEKAKANGCV 240
DB 181 QONGIGYVLNMSYTCPKPDFIPESHFLRPVNDSECEKILPWLIDKSVDFIEKAKANGCV 240
QY 241 LVHCLAGISRSATIIAIYMKRMDMSLDBATRFVKEKPTTSPNNPFGQLLDYEKKIKN 300
DB 241 LVHCLAGISRSATIIAIYMKRMDMSLDBATRFVKEKPTTSPNNPFGQLLDYEKKIKN 300
QY 301 QFGASGPKSKLKLHLKRNENEPVAVSFGQGSERPLSPPCADSATSEAGORPVHPASV 360
DB 301 QFGASGPKSKLKLHLKRNENEPVAVSFGQGSERPLSPPCADSATSEAGORPVHPASV 360
QY 361 PSVPSVQPSLLEDSPLVQALSGHLISADRLBDSNKLKRSFSLDIKSVSYSASMAASLHGF 420
DB 361 PSVPSVQPSLLEDSPLVQALSGHLISADRLBDSNKLKRSFSLDIKSVSYSASMAASLHGF 420
QY 421 SSSDALLEYKPSSTLLDGTNKLCOFSPVOELSEQTPETSPDKKASIPKKLQTARPSDSQ 480
DB 421 SSSDALLEYKPSSTLLDGTNKLCOFSPVOELSEQTPETSPDKKASIPKKLQTARPSDSQ 480
QY 481 SKRLHSVTRSSSGTQORSLISPLHRSQSVEDNYHTSPLFGLSTSQOHLTKSAGLKGWH 540
DB 481 SKRLHSVTRSSSGTQORSLISPLHRSQSVEDNYHTSPLFGLSTSQOHLTKSAGLKGWH 540
QY 541 SDILAPQSTPSTLSSWYFATESSHFYASAIYGSASYSAYSCSGLPTCGDQVYSVRRR 600
DB 541 SDILAPQSTPSTLSSWYFATESSHFYASAIYGSASYSAYSCSGLPTCGDQVYSVRRR 600
QY 601 QKPSDRAISRHWHSPEKQPKRRSCQMEFGESIMSENRREELGKVGSSSFGSME 660
DB 601 QKPSDRAISRHWHSPEKQPKRRSCQMEFGESIMSENRREELGKVGSSSFGSME 660
QY 661 IIEVS 665
DB 661 IIEVS 665
```

Search completed: June 21, 2004, 14:17:35
Job time : 60 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using bw model

Run on: June 21, 2004, 14:15:29 ; Search time 21 Seconds

(without alignments)
3046.065 Million cell updates/sec

Title: US-10-029-345A-109

Perfect score: 665
Sequence: 1 MAHEMIGTQIVTERLVALLE.....LGKVGSSQSSFGSGMERIEVS 665

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 473

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 75 summaries

Database : PIR 76:*

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match	Length	ID	Description

No matches found					

Search completed: June 21, 2004, 14:19:38
Job time : 21 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 14:11:38 ; Search time 17 Seconds
(without alignments)
2036.862 Million cell updates/sec

Title: US-10-029-345A-109
Perfect score: 665
Sequence: 1 MAHEMIGTQIVTERLVALLE.....LQKVGSQSSFSGSMEIIEVS 665

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 473

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 75 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query			
No.	Score	Match	Length	ID
Description				

No matches found

Search completed: June 21, 2004, 14:18:05
Job time : 17 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 21, 2004, 14:14:54 ; Search time 46 Seconds
(without alignments)
4561.236 Million cell updates/sec

Title: US-10-029-345A-109

Perfect score: 665

Sequence: 1 MAHEMIGTQIVTERIVALE.....LGKVGSSQSSFGSGMERIEVS 665

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 473

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 75 summaries

Database :

SPTREMBL_25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp Vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
No matches found					

Search completed: June 21, 2004, 14:19:04
Job time : 46 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: June 21, 2004, 14:17:40 ; Search time 49 Seconds
(without alignments)
3831.399 Million cell updates/sec

Title: US-10-029-345A-109
Perfect score: 665
Sequence: 1 MAHEMIGTQIVTERLVALL.....LGKVGSSQSSFSQSMETIEVS 665

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 1163542 seqs, 28233646 residues

Word size : 473

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 75 summaries

Database : Published Applications AA:*

```
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

No matches found					

Search completed: June 21, 2004, 14:21:16
Job time : 49 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: June 21, 2004, 14:16:30 ; Search time 23 Seconds
(without alignments)
1492.664 Million cell updates/sec

Title: US-10-029-345A-109

Perfect score: 665
Sequence: 1 MAHEMIGTQIVTERLVALL.....LGRVGSQSSFSGSMRIEVS 665

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 473

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 75 summaries

Database : Issued Patents AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
---------------	-------	----------------	--------	----	-------------

No matches found

Search completed: June 21, 2004, 14:20:14
Job time : 23 secs

This Page Blank (uspto)